Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Physics Reports 869 (2020) 1–51

Contents lists available at ScienceDirect

Physics Reports

journal homepage: www.elsevier.com/locate/physrep

COVID-19 and SARS-CoV-2. Modeling the present, looking at the future

Ernesto Estrada *

Instituto Universitario de Matemáticas y Aplicaciones, Universidad de Zaragoza, 50009 Zaragoza, Spain
ARAID Foundation, Government of Aragón, 50018 Zaragoza, Spain

ARTICLE INFO

Article history:
Received 27 July 2020
Accepted 27 July 2020
Available online 28 July 2020
Editor: I. Procaccia

ABSTRACT

Since December 2019 the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) has produced an outbreak of pulmonary disease which has soon become a global pandemic, known as COronaVIrus Disease-19 (COVID-19). The new coronavirus shares about 82% of its genome with the one which produced the 2003 outbreak (SARS CoV-1). Both coronaviruses also share the same cellular receptor, which is the angiotensin-converting enzyme 2 (ACE2) one. In spite of these similarities, the new coronavirus has expanded more widely, more faster and more lethally than the previous one. Many researchers across the disciplines have used diverse modeling tools to analyze the impact of this pandemic at global and local scales. This includes a wide range of approaches – deterministic, data-driven, stochastic, agent-based, and their combinations – to forecast the progression of the epidemic as well as the effects of non-pharmaceutical interventions to stop or mitigate its impact on the world population. The physical complexities of modern society need to be captured by these models. This includes the many ways of social contacts – (multiplex) social contact networks, (multilayers) transport systems, metapopulations, etc. – that may act as a framework for the virus propagation. But modeling not only plays a fundamental role in analyzing and forecasting epidemiological variables, but it also plays an important role in helping to find cures for the disease and in preventing contagion by means of new vaccines. The necessity for answering swiftly and effectively the questions: *could existing drugs work against SARS CoV-2?* and *can new vaccines be developed in time?* demands the use of physical modeling of proteins, protein-inhibitors interactions, virtual screening of drugs against virus targets, predicting immunogenicity of small peptides, modeling vaccinomics and vaccine design, to mention just a few. Here, we review these three main areas of modeling research against SARS CoV-2 and COVID-19: (1) epidemiology; (2) drug repurposing; and (3) vaccine design. Therefore, we compile the most relevant existing literature about modeling strategies against the virus to help modelers to navigate this fast-growing literature. We also keep an eye on future outbreaks, where the modelers can find the most relevant strategies used in an emergency situation as the current one to help in fighting future pandemics.

Contents

* Correspondence to: Instituto Universitario de Matemáticas y Aplicaciones, Universidad de Zaragoza, 50009 Zaragoza, Spain.

https://doi.org/10.1016/j.physrep.2020.07.005
0370-1573/

1. Introduction

In 2007, Cheng et al. [1] remarked that *"the presence of a large reservoir of SARS-CoV viruses in horseshoe bats, together with the culture of eating exotic mammals in southern China"* was *"a time bomb"*. The authors then warned that the possibility of the re-emergence of SARS and other new viruses from animals or laboratories should alert us about *"the need to be prepared"*. Twelve years later an epidemic outbreak expanded worldwide from its epicenter in Wuhan, province of Hubei, continental China, since December 2019. This new epidemic that eventually became a pandemic was named COVID-19. The **C**orona**v**irus **D**isease 20**19** (COVID-19) is a pulmonary disease produced by the **S**evere **A**cute **R**espiratory **S**yndrome **C**oronavirus **2** (SARS-CoV-2). The disease expanded quickly across the world from less than 30 cases at the end of December 2019 to more than 8,455,738 confirmed cases (according to existing data[1]) by June 20th, 2020 with the global evolution illustrated in Fig. 1.1.

The COVID-19 has expanded from its epicenter in the Hubei province to practically the whole world as can be seen in Fig. 1.2.[2] Since March 11th, 2020 it was declared a global pandemic by WHO, which is defined as 'the worldwide spread of a new disease' that can be spread worldwide because of the absence of preexisting immunity against the new pathogen in humans. By the end of January 2020 there were cases also reported from a number of countries that include: Taiwan, Thailand, Vietnam, Malaysia, Nepal, Sri Lanka, Cambodia, Japan, Singapore, Republic of Korea, United Arab Emirates, United States, The Philippines, India, Australia, Canada, Finland, France, and Germany. Five days after the declaration of the pandemic the disease had expanded to more than 140 countries.

According to the most recent scientific evidence, SARS CoV-2 is transmitted between people through respiratory droplets and contact routes (see [2]). The mechanism of virus transmission and some data about the virus is provided here according to data reported by Bar-On et al. [3], Rothan et al. [4], and Vannabouathong et al. [5]. Once the virus enters a human cell it has a *latent period* until the infected person becomes infectious himself. The latent period of SARS CoV-2 is approximately 3.69 days, which is then followed by an *infectious period* of about 3.48 days. When an infected individual is on the infectious period she can transmit the virus to other people by coughing or sneezing. Cough and sneeze produce droplets which can travel to another person with a proximity of about 2 m (see Fig. 1.3) who can have her mucosae or conjunctiva exposed to these droplets containing virion particles. Cough and sneeze produce droplets that travel at 10 m/s and 50 m/s, respectively. These respiratory droplets are formed of large particles ($> 5\ \mu$m) that fall quickly to the ground, and small droplets ($\leq 5\ \mu$m) which can evaporate to form 'droplet nuclei'. The large droplets may form an indirect route of transmission by indirect contact with surfaces in the immediate environment where they remain for periods ranging from 1 to 10 h. The droplet nuclei can remain suspended in air traveling to longer distances, and then could be inhaled by other individuals. Once entering an individual, the virus can infect several respiratory epithelial cells, namely type II pneumocytes, alveolar macrophages in the lung (see Fig. 1.3), and mucous glands in nasal cavity. The first symptoms appear after the *incubation period*, which has been estimated to be 5.1 days. Accordingly, almost 98% of those

1 https://ourworldindata.org/grapher/total-cases-covid-19?tab=map&year=2020-06-15.
2 https://ourworldindata.org/grapher/total-cases-covid-19?tab=map&year=2020-06-15.

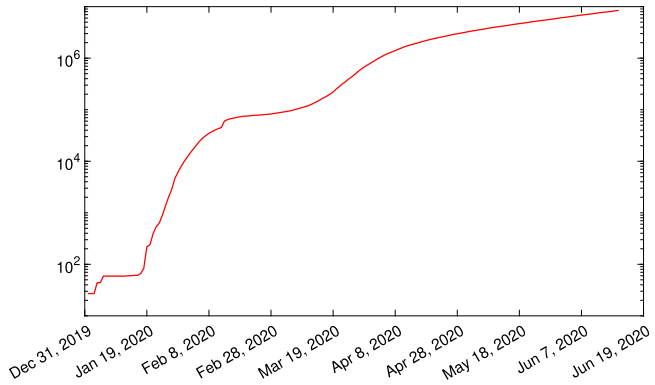


Fig. 1.1. Evolution of the number of confirmed cases in the world since December 31th 2019 until June 19th 2020. Notice the logarithmic scale on the y-axis displaying the number of confirmed cases.

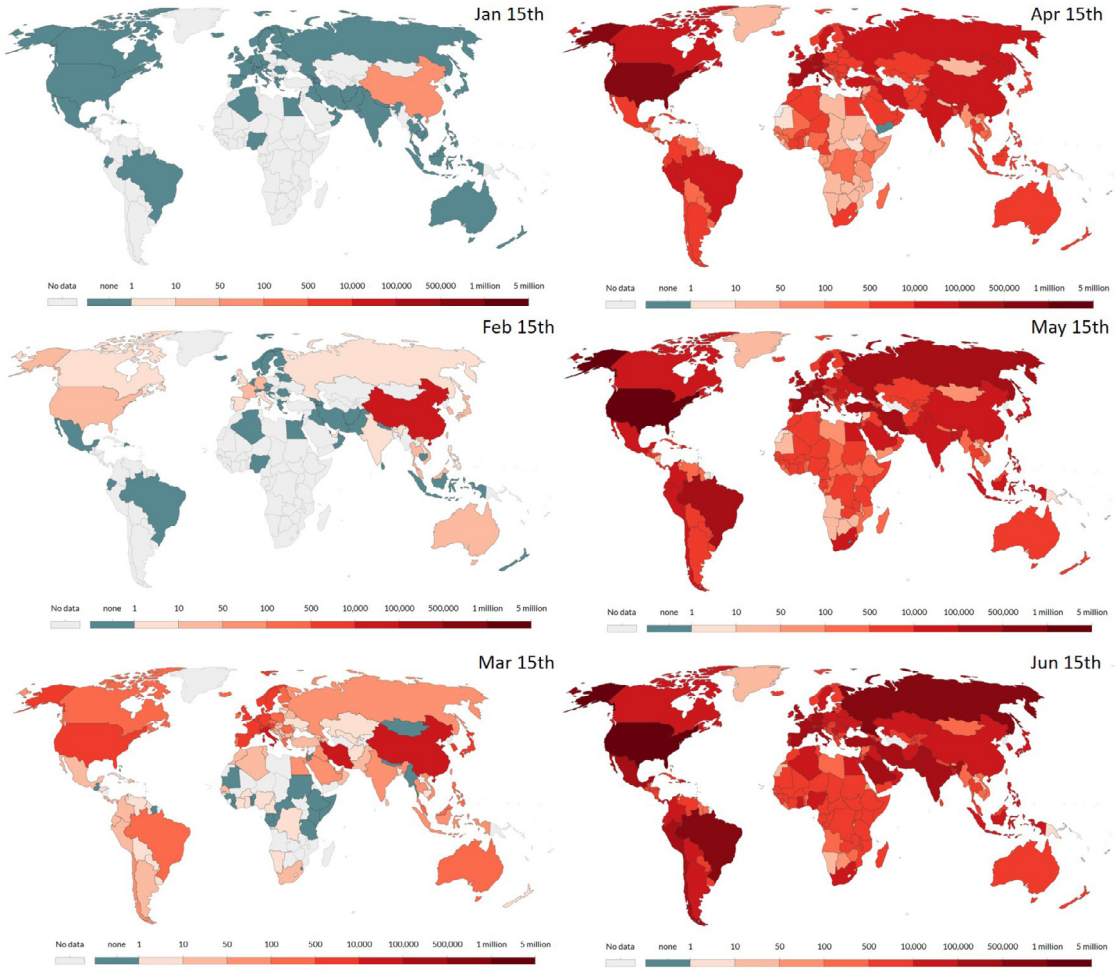


Fig. 1.2. Evolution of the COVID-19 pandemic across the world since January 2020 until Jun 2020.

infected individuals who will develop symptoms show them within 11.5 days after infection. These symptoms include fever, cough, and fatigue, while other symptoms include sputum production, headache, haemoptysis, diarrhea, dyspnoea, and lymphopenia. The main clinical diagnosis is pneumonia, but also other clinical manifestations include RNAemia, acute respiratory distress syndrome, acute cardiac injury, as well as incidence of grand-glass opacities that lead to death.

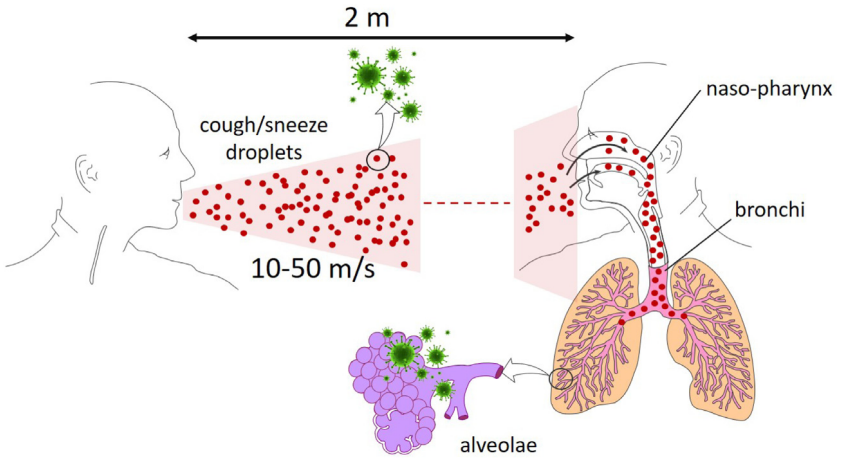


Fig. 1.3. Schematic representation of the main transmission route of SARS CoV-2 between individuals (see further for indirect transmission). The graphic was prepared using Motifolio.

Modeling and simulation have reached a great prominence during COVID-19 pandemic [6]. Kucharski [7] has remarked how models can help us not only to understand the past and present of an outbreak, but also to give clues about what might happen in the future by simulating many different alternatives. But modeling can not only help in understanding the epidemic evolution and planning different future scenarios, but it can also help in responding two other important questions about this epidemic: "*Could existing drugs work?*" and "*Can vaccines be developed in time?*" ([8]). In this review we focus on the three mentioned aspects more relevant for fighting COVID-19 pandemic: (i) the epidemiology of the pandemic, (ii) repurposing of known drugs to fight against the virus, and (iii) design of vaccines to protect against the virus. Therefore, the scope of this review is very broad and ambitious. It is broad because it starts by covering aspects of mathematical epidemiology, from simple compartmental models to more sophisticated ones, as well as models based on statistical and data-driven strategies. Then it continues with aspects related to drug repurposing, which include topics on biophysics, protein structure, drug-inhibitors interactions, and virtual screening of small molecules inhibiting pharmacological targets [9]. Finally, the review covers aspects related to vaccine design, including topics of vaccinomics and immunoinformatics, and making emphasis on the design of epitope vaccines. To make this review comprehensive to a wide audience it includes aspects of the virology of SARS Co-2, its transmission, life cycle, and its entry into human cells, all based on the most recent findings from structural biology and immunology. The review is written in the middle of the global pandemic hitting most countries in world, and in a situation in which new information about the epidemiology, virus biology and development of drugs/vaccines advance in a per day way.

2. Epidemiological analysis

Epidemiological analysis is a well-established area of interdisciplinary research in which medical, statistical, computational and mathematical elements are integrated to analyze and forecast the progression of a disease. In the words of Vespignani [10] we should differentiate these studies in 'peace time', when there is no health emergency, from those of 'war times', when the emergency of a pandemic is pressing for results with limited data, in a constantly changing landscape and having to make a lot of assumptions. Tian and Dye (in [10]) have also stressed that in this scenario we cannot test the effectiveness of control measures in an experiment or a clinical trial, but instead we have to rely on statistical and mathematical modeling. Therefore, in this section we reviewed some of the most important models and findings reported in the literature so far for the analysis and forecast of the COVID-19 pandemic. We should remark that as stressed by Vespignani [10]: "*what has been produced the day before often must be completely revised the day after because a new piece of information has arrived*".

2.1. Basic compartment models

The deterministic mathematical models used to analyze the coronavirus rest on the shoulders of a pair of giants: A.G. McKendrick and W.O. Kermack, the first military doctor and the second chemist. They developed the first compartmental mathematical model for the study of the evolution of an epidemic [11]. In compartmental models the population is divided into different compartments, such as Susceptible (S), Infectious (I) and Recovered (R). The differential equations are then written to analyze the progression of the compartments taking into account the probability that an I–S contact will produce a new infected person, or the probability that an infected person will recover and go from the compartment I to the compartment R. These probabilities are usually estimated from the empirical data of an infection and are vital for simulations of its progression. Here we introduce the basic concepts of these compartment models which will be necessary for the understanding of more complex models used to forecast the evolution of the COVID-19 pandemic.

2.1.1. Susceptible–Infected–Recovered (SIR) model

Let β be the infection rate and let $I(t)$ and $S(t)$ be the fractions of infected and susceptible individuals at time t. Let γ be the rate at which infected individuals recover, and let $R(t)$ be the fractions of recovered individuals. Then the Susceptible–Infected–Recovered model has the following scheme and scalar equations:

$$\begin{aligned} \dot{S}(t) &= -\beta S(t) I(t), \\ \dot{I}(t) &= \beta S(t) I(t) - \gamma I(t), \\ \dot{R}(t) &= \gamma I(t). \end{aligned} \tag{2.1}$$

An important parameter in mathematical epidemiology, which has also been widely studied for COVID-19 (see [12]), is the basic reproductive number, which is defined below.

The *basic reproductive number* R_0 is the average number of new infections caused by individuals who are infected shortly after disease introduction in a completely susceptible population. If $R_0 > 1$ the disease can propagate and become an epidemic, while if $R_0 < 1$, the disease will die out.

The assumption that the whole population is susceptible is very strong, as there could be contacts which are immune. Therefore, not all contacts of an infectious individual are susceptible to get the virus and we need a new parameter that takes into account this proportion of the population which is susceptible to a disease. This parameter is defined below.

The *effective reproductive number* $R_e(t)$ is the number of new infections caused by a single infectious individual at time t in a partially susceptible population. Then,

$$R_e(t) = R_0 S(t). \tag{2.2}$$

Let us consider that $S(0) > 0, I(0) > 0$ and $S(t) + I(t) + R(t) = 1$. If $\beta S(0)/\gamma > 1$ then $t \mapsto I(t)$ increases first monotonically to a maximum value

$$I_{max} = I(0) + S(0) - \frac{\gamma}{\beta}\left(\log(S(0)) + 1 - \log\left(\frac{\gamma}{\beta}\right)\right), \tag{2.3}$$

and then decreases monotonically to 0 as the time growth to infinity. Then, the basic and effective reproductive numbers in this model are given by $R_0 = \beta/\gamma$ and $R_e(t) = \beta S(t)/\gamma$, respectively.

It should be remarked here that in real-life examples contact rates and transmissibility typically change over time, which is specially true after contention measures are observed. In this case, a better epidemiological parameter to capture these realities is the case (also cohort) reproduction number at time step t, R_t^c. It counts the average number of secondary transmissions caused by a cohort infected at time step t. As pointed out by Cori et al. [13] R_t^c can be calculated a posteriori, once the secondary cases generated by cases infected at t have been infected.

An epidemiological model can also be studied on a network representing the interactions between individuals (contact network), or representing the mobility between regions or patches. In general a network is a weighted graph (see Fig. 2.1 (left)) $G = (V, E, W, \varphi)$, where a node $i \in V$ represents an individual, institution, geographic region, and so forth, and two nodes i and j form a directed edge $(i, j) \in E$ if there is a "flow" from i to j. $W = \{w_{ij} \in \mathbb{R}\}$ is a set of weights assigned to the edges by the function $\varphi : (i, j) \to w_{ij}$ which may represent a probability of transition, a density of flow between the nodes or the strength of a social tie. A self-loop is an edge $(i, i) \in E$. A graph in which $(i, i) \notin E$ for all $i \in V$, and for which $(i, j) \in E$ implies that $(j, i) \in E$ with $w_{ij} = w_{ji} = 1$ is a simple graph or network. A multilayer network (2.1 (right)) is a graph $G = (\cup_i V_i, \cup_i E_i)$ where the subsets of vertices V_i may represent entities of one class different from those in the set V_j, for instance trains and cars [14,15]. Each of these sets of nodes is then represented as a layer of the multilayer graph. There are intra- and inter-layer edges as illustrated in Fig. 2.1 (right). The adjacency matrix A of a weighted directed graph is a square matrix whose entries $A_{ij} = w_{ij}$ for every pair of (not necessarily different) vertices $i, j \in V$. In case of simple graph, A is symmetric with $A_{ij} = 1$ if $(i, j) \in E$ and $A_{ij} = 0$ otherwise.

In a network of interactions the SIR equations are transformed to [16]:

$$\begin{aligned} \dot{S}_i(t) &= -\beta S_i(t) \sum_{j=1}^{n} A_{ij} I_j(t), \\ \dot{I}_i(t) &= \beta S_i(t) \sum_{j=1}^{n} A_{ij} I_j(t) - \gamma I_i(t), \\ \dot{R}_i(t) &= \gamma I_i(t). \end{aligned} \tag{2.4}$$

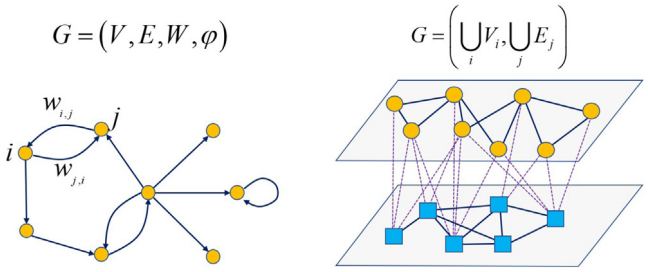


Fig. 2.1. Illustration of a weighted graph (left) and a multilayer graph (right).

In this case the set of equilibrium points is the set of pairs $\left(S^*, \vec{0}\right)$, for any $S^* \in [0, 1]^n$. The linearization of the network SIR model about the equilibrium point $\left(S^*, \vec{0}\right)$ is:

$$\begin{aligned} \dot{S}(t) &= -\beta \text{diag}\left(S^*\right) AI(t), \\ \dot{I}(t) &= \beta \text{diag}\left(S^*\right) AI(t) - \gamma I(t). \end{aligned} \tag{2.5}$$

Let $G(t) = \text{diag}(S(t))A$ with eigenvalues $\lambda_1(t) > \lambda_2(t) \geq \cdots \geq \lambda_n(t)$ and let $\psi_j(t)$ be the eigenvector associated with the jth eigenvalue of $G(t)$. Then, following Mei et al. [16] if we multiply the expression for $\dot{I}(t)$ in Eq. (2.5) by $\psi_1^T(\tau)$ from the left, we get:

$$\begin{aligned} \frac{d}{dt}\left(\psi_1^T(\tau) I(t)\right) &= \psi_1^T(\tau)\left(\beta \text{diag}(S(t)) AI(t) - \gamma I(t)\right) \\ &\leq \psi_1^T(\tau)\left(\beta \text{diag}(S(\tau)) AI(t) - \gamma I(t)\right) \\ &= \left(\beta\lambda_1(\tau) - \gamma\right)\psi_1^T(\tau) I(t). \end{aligned} \tag{2.6}$$

Thus,

$$\psi_1^T(\tau) I(t) \leq \left(\psi_1^T(\tau) I(0)\right) e^{(\beta\lambda_1(\tau)-\gamma)t}. \tag{2.7}$$

Because the exponential dies out when $\beta\lambda_1(\tau) < \gamma$ we have that $\psi_1^T(\tau) I(t)$ monotonically decays to zero for all $t > \tau$, which indicates that when $\beta\lambda_1(\tau) < \gamma$ the epidemic dies out. Now, applying a similar strategy but using $\psi_1^T(0)$ we have

$$\begin{aligned} \left.\frac{d}{dt}\left(\psi_1^T(0) I(t)\right)\right|_{t=0} &= \psi_1^T(0)\left(\beta \text{diag}(S(t)) AI(t) - \gamma I(t)\right) \\ &= \left(\beta\lambda_1(0) - \gamma\right)\psi_1^T(0) I(t), \\ &> 0. \end{aligned} \tag{2.8}$$

Then, for small t the weighted average $t \mapsto \psi_1^T(0) I(t)$ grows exponentially fast with rate $\beta\lambda_1(0) - \gamma$. Additionally, it can be proved that there exists $\tau > 0$ such that $\beta\lambda_1(0) < \gamma$. Therefore, the effective reproductive number for the network SIR model is $R_e(t) = \beta\,\lambda_1(t)/\gamma$. If we consider that very close to $t = 0$ all individuals are susceptible, i.e., $S(0) = \vec{1}$ (where $\vec{1}$ is the all-ones vector), then $R_0 = \beta\,\lambda_1(A)/\gamma$, where $\lambda_1(A)$ is the spectral radius of the adjacency matrix [16].

Although the SIR model is very simple and does not capture all the compartments in which a population is divided in a realistic COVID-19 situation, it has been used for the prediction of the evolution of this epidemic. In one of these works D'Arienzo and Coniglio [17] studied the values of R_0 for SARS-CoV-2 using data derived from the early phase of the outbreak in Italy. They evaluated the basic reproductive number in 9 Italian cities using the SIR model in the interval between February 2 and March 12, 2020. Their estimation of R_0 for the whole country ranges between 2.43 and 3.10 with the highest value for the city of Lodi (3.09) followed by Cremona (2.76). Wangping et al. [18] also used SIR to estimate the basic reproductive number in Italy from January 22th, 2020 to April 2th, 2020, and compare it with that of the city of Hunan in China. They reported a mean value of 3.10 for R_0 in Italy and of 2.48 for Hunan. These values are elevated to 4.34 and 3.16, respectively by considering an extension of the SIR model (eSIR) which includes the effects of different intervention measures in dissimilar periods. According to the eSIR model "there would be a total of 182,051 infected cases (95%CI:116 114–274,378) under the current country blockade and the endpoint would be Aug 05 in Italy". By June 21th 2020 the number of COVID confirmed cases in Italy was 238,275, well over those predicted by the eSIR model.

The estimation of R_0 was also the goal of You et al. [19] for several cities in China. They also used the SIR model and compared it with the prediction made by Poisson likelihood-based method (ML), and the exponential growth rate-based method (EGR). They considered data until March 31, 2020. In all cases the values of R_0 estimated with SIR were significantly larger than those estimated with ML and EGR for all cities. They also considered the controlled reproduction

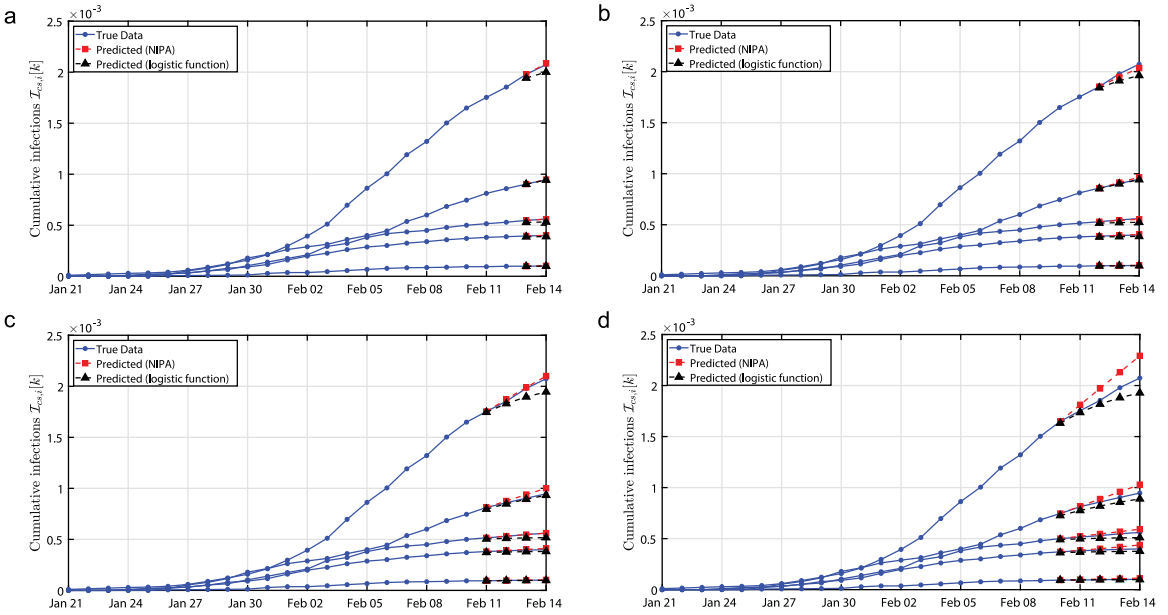


Fig. 2.2. Prediction made by Prasse et al. [23] for the outbreak in Hubei for 4 cities. Reproduced with permission by the authors.

number R_c, which describes the ability of disease spreading when interventions are taking place. Thus it takes into account the effects of measures such as quarantine, isolation, or traffic control. In all cities analyzed in this paper $R_c < R_0$ by the three approaches used in the study. In this case, SIR approach does not give the highest values of the estimation like for the case of R_0, but in general it gives values close to those predicted by EGR.

An interesting paper was published by Roda et al. [20] with the suggestive title "Why is it difficult to accurately predict the COVID-19 epidemic?" The authors compare the predictions made by the simple SIR model with the SEIR one (see next section) for the COVID-19 in Wuhan city based on the official reports from January 21 to February 4, 2020. They used the Akaike Information Criterion to compare the standard SIR and SEIR models in predicting the epidemic and concluded that "given the same dataset of confirmed cases, more complex models may not necessarily be more reliable in making predictions due to the larger number of model parameters to be estimated". Based on the SIR model Roda et al. [20] demonstrated the existence of a linkage between the transmission rate and the case-infection ratio, which significantly affects the predictions of the epidemic made by the model. According to these authors this is the cause for variability in model predictions and it should not be interpreted "as a shortcoming of transmission models; neither is it caused by the limited number of time points in data. Rather, it is caused by the lack of datasets that are independent of the confirmed cases to allow modelers to produce independent estimates of" both parameters mentioned before.

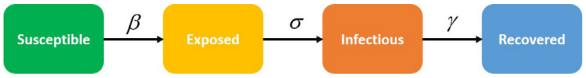
A small variation of the SIR model, known as SIRD, has also been used for the study of COVID-19. In this model the population is subdivided into the susceptible (S), infected (I), recovered (R), and dead (D) compartments. Fanelli and Piazza [21] used this model to analyze the temporal dynamics of the COVID-19 outbreak in China, Italy and France in the time window 22th January to 15th March 2020. According to their predictions the peak in Italy should occurs around March 21st 2020, with a peak number of infected individuals of about 26,000 (not including recovered and dead) and a number of deaths at the end of the epidemic of about 18,000. Remarkably, according to WHO data, the peak in Italy occurred on March 21st. However the number of death in Italy by June 21th is already 34,634. The SIRD model was also used by Anastassopoulou et al. [22] for the analysis of the epidemic in Hubei, China from January 11 to February 10, 2020. They estimated the case fatality and case recovery ratios, along with their 90% confidence intervals as the outbreak evolves, as well as the basic reproduction number, and the per day infection mortality and recovery rates.

A network based SIR approach was developed by Prasse et al. [23] to describe the COVID-19 epidemic in the Hubei province. They considered a network of 16 cities in Hubei. In this case the infection rate β is disaggregated into inter-cities infection rate β_{ij}, which are the infection probability from city j to city i. This gives rise to a weighted adjacency matrix in which the entries correspond to the values of β_{ij} for a given pair of cities. Such probabilities can be obtained from mobility patterns between cities, but in this case they were inferred from observing the evolution of the epidemic. In Fig. 2.2 we reproduce the results obtained by Prasse et al. [23] for 4 of the 16 cities studied.

2.1.2. Susceptible–Exposed–Infected–Recovered (SEIR) model

As we have explained in Section 2, SARS CoV-2 has a latent period in which an individual is infected but is not yet infectious. In compartment models, this period is named Exposed (E) and it is an intermediate period between being

susceptible (S) and becoming infectious (I). The resulting model is a four compartment one known as SEIR, whose scheme and equations for the scalar model are given below:



$$\begin{aligned}
\dot{S}(t) &= -\beta S(t) I(t), \\
\dot{E}(t) &= \beta S(t) I(t) - \sigma E(t), \\
\dot{I}(t) &= \sigma E(t) - \gamma I(t) \\
\dot{R}(t) &= \gamma I(t).
\end{aligned} \tag{2.9}$$

Based on the results of Wallinga and Lipsitch [24], Zhou et al. [25] and Fang et al. [26] have reported the values of the basic reproductive number using

$$R_e \approx \left(1 + \frac{\lambda}{\sigma}\right)\left(1 + \frac{\lambda}{\gamma}\right), \tag{2.10}$$

where $\lambda = \ln Y(t)/t$ and $Y(t)$ is the number of infected individuals with symptoms at time t. Zhou et al. [25] estimated R_0 to fall between 2.8 and 3.3 by using the real-time reports on the number of COVID-19 infected cases from People's Daily in China and fall between 3.2 and 3.9 on the basis of the predicted number of infected cases from international colleagues. Fang et al. [26] estimated the values of $R_0(t)$ showing values in the range between 2.47 on January 20th to 2.34 on February 29th, with a maximum of 3.20 on January 30th. A simple SEIR model assuming well-mixed population was used by Hou et al. [27] for exploring the effectiveness of the quarantine of the Wuhan city against this epidemic, on the transmission dynamics of COVID-19. They show that by reducing the contact rate of latent individuals, interventions such as quarantine and isolation, can effectively reduce the potential peak number of COVID-19 infections and delay the time of peak infection.

An important application of SEIR model was developed by Fox et al. [28] who used it to model the number of hospitalized cases, and Intensive Care Unit (ICU) cases, per 100,000 population. They considered two scenarios, one in which there was no intervention which displays $R_0 = 2.4$ and a second one with social interventions dropping the basic reproductive number to $R_0 = 1.6$, both with a start prevalence on 1st March 2020 of 2 persons per million. The authors compared the results of the SEIR model with those obtained by a model developed by Ferguson et al. [29] on the basis of a parameterized individual-based simulation model. According to [28] the peak demand for ICU beds was estimated to be 6965 according to Ferguson et al. [29] model or 5109 ICU beds according to SEIR.

Due to the importance of R_0 in all modeling scenarios of COVID-19 we mention here a result reported by Grant [30]. Due to the fact that compartment models do not accurately capture the distribution of times that an individual spends in each compartment, they do not accurately capture the transient dynamics of epidemic. In the case of SEIR, Grant [30] shows that the model overestimates epidemic duration by a factor of 2. It also underestimates peak infection rates by a factor of 3 based on the progress of the epidemic in Wuhan.

In the case in which a network of interactions is taken into account, the system can be expressed in matrix–vector form as:

$$\begin{aligned}
\dot{\vec{S}}(t) &= -\beta \mathrm{diag}\left(\vec{S}(t)\right) A\vec{I}(t), \\
\dot{\vec{E}}(t) &= \beta \mathrm{diag}\left(\vec{S}(t)\right) A\vec{I}(t) - \sigma \vec{E}(t) \\
\dot{\vec{I}}(t) &= \sigma \vec{E}(t) - \gamma \vec{I}(t).
\end{aligned} \tag{2.11}$$

The importance of including information about the contact networks to analyze the evolution of (COVID-19) epidemic was stressed by Small and Cavanagh [31]. According to their results the knowledge of the topology of these contact networks is more important than the precise knowledge of epidemic transmission parameters. Their model consists of city-level transmission of an infectious agent with a SEIR dynamic. They then modeled the topological patterns of contact by varying the structure of the network between different standard topologies, such as: scale-free contact network, a random graph, regular lattice, and small world networks. Their model exhibits good agreement between simulation and data from the 2020 pandemic spread of COVID-19 in several Australian cities.

One of the networks that can be used in these analyses is transportation network, such as the network of air traffic between different airports. This strategy was followed by Linka et al. [32] by using the air transportation network of Europe in combination with a SEIR model. They correlated the mobility model to passenger air travel statistics and calibrated the SEIR model using the number of reported COVID-19 cases for each country in Europe. According to their results mobility networks of air travel can predict the emerging global diffusion pattern of a pandemic at the early stages of the outbreak. In addition they compared the propagation of the disease with unconstrained mobility and by implementing reduced

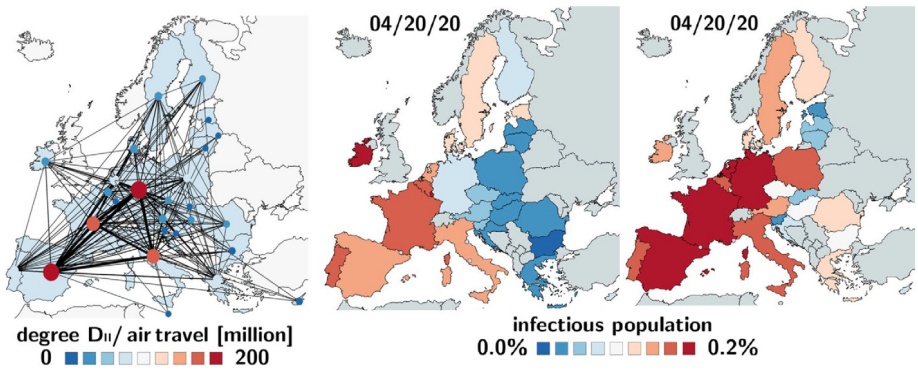


Fig. 2.3. Illustration of the mobility network of the European Union (left), and of the effects of travel restrictions according to Linka et al. [32]. Figures provided by the authors.

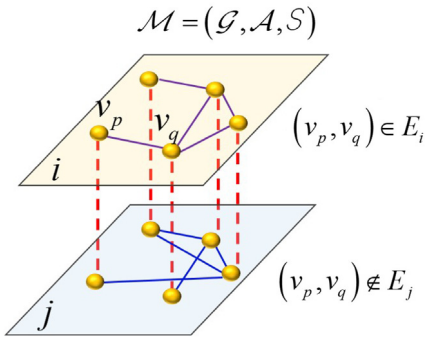


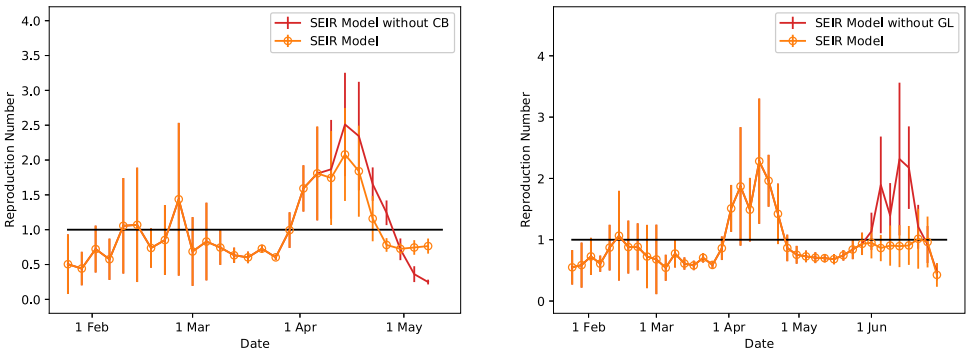
Fig. 2.4. Illustration of a multiplex formed by two layers i and j. Notice that the set of vertices is the same in each of the two graphs G_i and G_j, but not the sets of edges $E_i \neq E_j$.

mobility strategies, clearly showing that unconstrained mobility would accelerate significantly the propagation across Europe (see Fig. 2.3).

In order to grasp the complexities of the contact networks existing in modern society and necessary for modeling the propagation of a disease in an effective way it is sometimes needed to use a more complex representation. A condensed representation of different facets of the same set of entities, e.g., the different contact ways between the same group of people, can be obtained by means of a multiplex [14,15], which is the triple $\mathcal{M}=(\mathcal{G},\mathscr{A},S)$, where $\mathcal{G}=\{G_1,\ldots,G_h\}$ is a set of weighted graphs, in which $G_i=(V,E_i)$ where $V=\{v_1,\ldots,v_n\}$ is a set of nodes and $E_i=\left\{\left(v_p\left(i\right),v_q\left(i\right)\right)\right\}$ is a set of edges between these nodes in the layer i, $\mathscr{A}=\{\mathscr{A}(v_1),\ldots,\mathscr{A}(v_n)\}$ is a set of node identities, such that the following equivalence relation exists: $v_p\left(G_1\right)\overset{\mathscr{A}}{\equiv}v_p\left(G_2\right)\overset{\mathscr{A}}{\equiv}\cdots\overset{\mathscr{A}}{\equiv}v_p\left(G_h\right)$ for every node $v_p(i)$. Finally, $S=\left\{S_{1,2},\ldots,S_{h-1,h}\right\}$, where $S_{ij}=\left\{\omega_1^{i,j},\ldots,\omega_n^{i,j}\right\}$ is the set of interlayer links between the nodes and their copies in different layers. The term $\omega_p^{i,j}=\omega\left(v_p\left(i\right),v_p\left(j\right)\right)$ represents a weight (known as coupling strength) for the pair $v_p\in V_i$, $v_p\in V_j$ for $i\neq j$. An illustration is given in Fig. 2.4.

Chung and Chew [33] considered a multiplex network consisting of 5 layers: (i) a household network, (ii) a dormitory network, (iii) a workplace network, (iv) a temporal crowd network, and (v) a temporal social gathering network. The household network accounts for the interactions among family members within a household. It is modeled as a network of communities of nodes densely connected among them (households) but poorly connected among them. The dormitory network describes the interactions between workers in dormitory houses, which hold much more people than houses. The network of workplace interactions accounts for the social contacts among workers at workplace, which in Singapore represents 40% of the total population. The temporal crowd network accounts for interactions by random groups of people at public spaces, such as public transportation system and markets which have short-term duration each day. Finally, the layer of social gathering accounts for events such as religious services, academic conferences, and large-scale dinner events. The authors focused on the analysis of the epidemic progression in Singapore starting on January 21st, 2020, two days before the first reported case of COVID-19 in Singapore, and ending on May 13th, 2020. Singapore implemented contention measures (Circuit Breaker, CB) on April 7th, 2020 to restrict social interactions among its residents.

The multiplex SEIR model was used to analyze the evolution of R_0 (namely it should be said $R_e\left(t\right)$) for the period of analysis considering a population of 1 million nodes in the multiplex. In the evolution of this parameter until May 1st



(a) (b)

Fig. 2.5. Illustration of the results obtained by Chung and Chew [33] for the dynamical evolution of the reproduction number with and without circuit breaker (CB) (a), and with and without gradual lifting (G) (b) measures. Figures provided by the authors.

it can be seen as two small waves followed by a bigger one (see Fig. 2.5). The first occurred on 9 Feb 2020 produced by an infection cluster at the church of the Grace Assembly of God. The second was on 23 Feb 2020 and it is related to a private function at SAFRA Jurong. The biggest wave occurred on 5 April 2020 due to reported cases of infection within the construction site in the Raes Place area which involved 23,000 foreign workers in their dormitories. Chung and Chew [33] also used their model to forecast the effects of Gradual Lifting (GL) of contention measures at the beginning of June and the prediction of big wave in case no such gradual lifting is taken into account (see Fig. 2.5).

A work that combined the SEIR model with statistical inference of parameters from linear regression was developed by Tian et al. [34] who investigated the spread and control of COVID-19 during the first 50 days of the epidemic in China. They used data about case reports, human movement, and public health interventions. Using linear regression they found associations between the transmission control measures and the number of cases reported during the first week of an outbreak in a new location. They found that the Wuhan lockdown is associated with the delayed arrival of COVID-19 in other cities by 2.91 days. Cities that implemented control measures preemptively reported fewer cases on average (13.0) in the first week of their outbreaks compared with cities that started control later (20.6). As expected, banning intracity public transport, closing entertainment venues, and closing public gatherings were associated with reductions in case incidence. Another work combining SEIR and data-driven analysis was performed by Manchein et al. [35] to analyze the growth of the cumulative number of confirmed infected cases by COVID-19. They found that: (i) power-law growth is observed in all countries; (ii) by using the distance correlation, the power-law curves between countries are statistically highly correlated. This suggested a universality of such curves around the world; and (iii) soft quarantine strategies are inefficient to flatten the growth curves. They also reported that besides the social distancing of individuals the identification and isolation of infected individuals in a large daily rate can help to flatten the power-laws.

A SEIR model combined with network analysis was also used by Peirlinck et al. [36] to study the outbreak of COVID-19 in China and USA. They found a latent period of 2.56 $\pm$0.72 days for the outbreak in China with a contact period of 1.47 $\pm$0.32 days, and an infectious period of 17.82 $\pm$2.95 days. In the USA the contact period estimated is of 3.38 $\pm$ 0.69 days, and a basic reproduction number of 5.30 $\pm$ 0.95, which seems very high in comparison with other reports. They then estimated a USA nationwide peak of the outbreak on May 10th, 2020 with 3 million infections. As on 3rd July 2020 the number of infected individuals is 2.4 million, which is still well below the estimation for the May 10th, and the number of cases is still growing, which indicates that the peak has not been reached yet.

Kucharski et al. [37] combined a modified SEIR model with data analysis for the study of the epidemic in Wuhan and international cases that originated in Wuhan. They estimated how transmission had varied over time during January, 2020, and February, 2020. They used data about daily number of new internationally exported cases (or lack thereof), by date of onset, as of Jan 26, 2020; daily number of new cases in Wuhan with no market exposure, by date of onset, between Dec 1, 2019, and Jan 1, 2020; daily number of new cases in China, by date of onset, between Dec 29, 2019, and Jan 23, 2020; and proportion of infected passengers on evacuation flights between Jan 29, 2020, and Feb 4, 2020. In addition, the authors also considered the daily number of new exported cases from Wuhan (or lack thereof) in countries with high connectivity to Wuhan (i.e., top 20 most at-risk countries), by date of confirmation, as of Feb 10, 2020; and data on new confirmed cases reported in Wuhan between Jan 16, 2020, and Feb 11, 2020. This work reports a daily reproductive number of 2.35 one week before travel restrictions and a drop to 1.05 one week after it. They calculated that in locations with similar transmission potential to Wuhan in early January, a 50% chance of infection is obtained with at least four independently introduced cases.

In another approach, Block et al. [38] combined a SEIR model with the following two ingredients: (i) a network of contact probabilities in the population representing the typical contact people had in a pre-COVID-19 world. They considered ties between individuals who live geographically close, individuals who are similar on individual

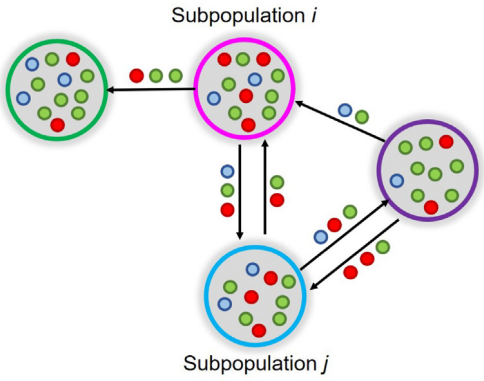


Fig. 2.6. Illustration of a metapopulation network.

attributes – age, education or income – and individuals who are members of common groups, such as households and institutions (including schools and workplaces). Additionally, random connections in the population were added; (ii) actors interact at discrete times with others from their personal network. They then focus on some possible strategies to reduce the COVID-19 spreading in a post-lockdown world. These strategies are: (a) limiting interaction to a few repeated contacts akin to forming social bubbles; (b) seeking similarity across contacts; and (c) strengthening communities via triadic strategies. According to their simulation results "a strategic social network-based reduction of contact strongly enhances the effectiveness of social distancing measures while keeping risks lower".

The combination of a simple SEIR model and contact tracing allowed Keeling et al. [39] to predict that under effective contact tracing less than 1 in 6 cases will generate any subsequent untraced infections. This intensive tracing will come at a high logistical burden with an average of 36 individuals traced per case. Changes to the definition of a close contact can reduce this burden, but with increased risk of untraced cases. If such definition is too relaxed, then it is unlikely to control spread. The data for this study was generated by the authors by reporting 50,000 encounters between 5802 respondents of a cross-sectional survey. The authors then extrapolated this data to generate a pattern of contacts over a 14-day period. Then, using a latent period of the disease of 4 days, a basic reproductive number of 3, the results were obtained from the SEIR simulations on the network of contacts.

Another variation of interconnected system which is common in epidemiological modeling is the use of metapopulation networks. A metapopulation network is a graph $G = (V, E)$, where every node $v_p \in V$ is formed by $N_p \in \mathbb{N}$ particles (individuals), which are then colored by a fixed number of colors $C \in \mathbb{N}$, which represent the states of the individuals, e.g., Susceptible, Infected, Recovered. The directed edges $\left(v_p, v_q\right)$ connects the corresponding nodes in such a way that they moves particles of different colors between one node to another as illustrated in Fig. 2.6.

Aleta and Moreno [40] implemented a data-driven version of a metapopulation SEIR model that allows obtaining realistic estimates for the spatial incidence of the disease as well as its temporal dynamics. They applied their model to the epidemic spreading in Spain, where the spatial network considered consisted of the 52 Spanish provinces. Then, the disease dynamics is modeled by combining the SEIR model and the mobility of the individuals across the subpopulations. Interestingly, the authors used a breakdown of inter-province flows in Spain by transportation mode, which includes airplane, coach, car, ship and train. Due to the differences in the communication routes and geographic position of a given region respect to another, under the same conditions, the spatial spreading of the disease would be completely different if the epidemic starts in one region or another. This is well-illustrated by the cases of Barcelona and Madrid, which are the most important cities in Spain (see Fig. 2.7).

2.2. Models with increased number of compartments

In search of a more realistic description of the COVID-19 outbreak, the number of compartments in which the population is subdivided needs to be increased. This is the case for the consideration of the existence of an environmental reservoir. As described in Section 2 the virus contained in large droplets expelled by an infectious individual can remain at surfaces for periods of time that allow susceptible individuals to be in contact with it as illustrated in Fig. 2.8.

This environmental reservoir (V) can be added as a new compartment into a SEIR model such that we have the following SEIRV model [41]:

$$\begin{aligned}
\dot{S}(t) &= \Lambda - \beta_E(E) S(t) E(t) - \beta_I(I) S(t) I(t) - \beta_V(V) S(t) V(t) - \mu S(t), \\
\dot{E}(t) &= \beta_E(E) S(t) E(t) + \beta_I(I) S(t) I(t) + \beta_V(V) S(t) V(t) - (\alpha + \mu) E(t), \\
\dot{I}(t) &= \alpha E(t) - (w + \gamma + \mu) I(t) \\
\dot{R}(t) &= \gamma I(t) - \mu R(t) \\
\dot{V}(t) &= \xi_1 E(t) + \xi_2 I(t) - \sigma V(t),
\end{aligned} \tag{2.12}$$

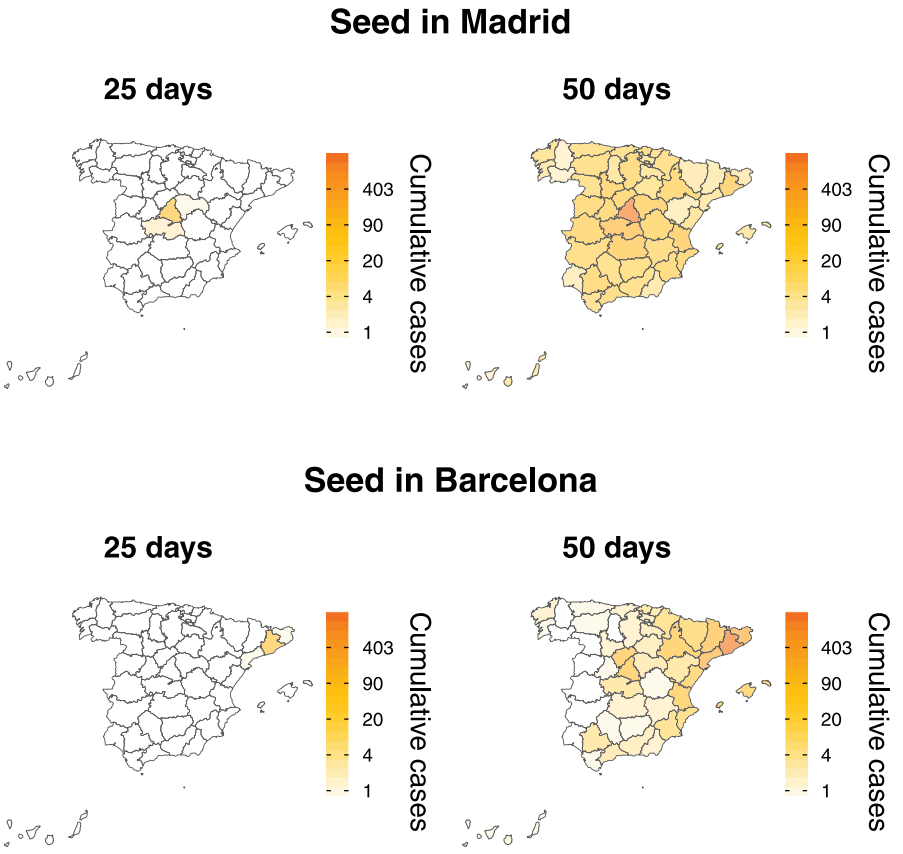


Fig. 2.7. Comparison of the spreading from the two major cities of Spain at to different times as reported by Aleta and Moreno [40]. The reported values are the median over 103 simulations. The figure is reproduced with permission from the authors.

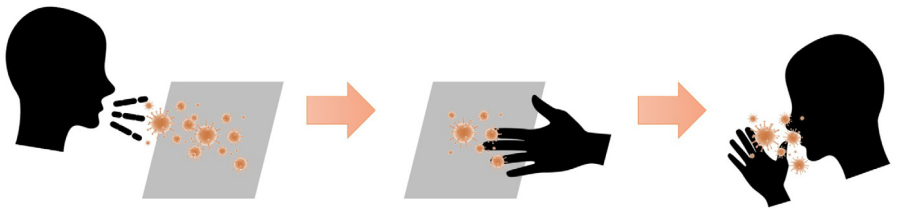


Fig. 2.8. Schematic illustration of the indirect transmission of COVID-19. An individual sneeze or cough over a surface (left), which is then touched by another individual (center) who brings the virus from her hands to her respiratory system (right).

where,

V: concentration of the coronavirus in the environmental reservoir;
Λ: population influx;
$\beta_E\,(E)$: direct human-to-human transmission rate between the exposed and susceptible individuals;
$\beta_I\,(I)$: transmission rate between the infected and susceptible individuals;
$\beta_V\,(V)$: indirect environment-to human transmission rate;
μ: natural death rate of human hosts;
α^{-1}: incubation period between the infection and the onset of symptoms;
γ: rate of recovery from infection;
ξ_1: rate of the exposed individuals contributing the coronavirus to the environmental reservoir;
ξ_2: rate of the infected individuals contributing the coronavirus to the environmental reservoir;
σ: removal rate of the virus from the environment.

This is the model that Yang and Wang [41] have developed to describe the COVID-19 epidemic in Wuhan, China. They have assumed that $\beta_E\,(E)$, $\beta_I\,(I)$ and $\beta_V\,(V)$ are all non-increasing functions and for performing the simulations they

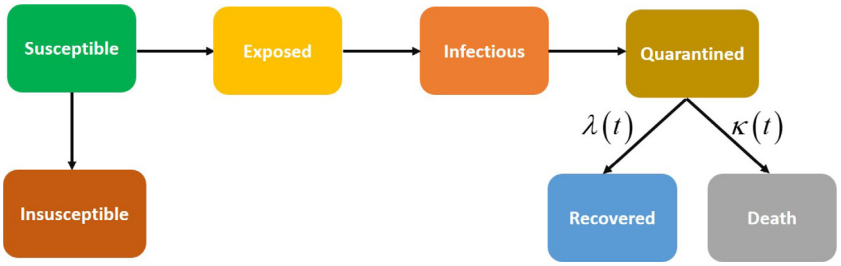


Fig. 2.9. Scheme of the compartments used in the model developed by Godio et al. [44].

considered that the three transmission rates can be written as:

$$\beta_E(E) = \frac{\beta_{E0}}{1+cE},\ \beta_I(I) = \frac{\beta_{I0}}{1+cI},\ \beta_V(V) = \frac{\beta_{V0}}{1+cV}, \tag{2.13}$$

where $c > 0$ is a coefficient. Yang and Wang [41] estimated mean values for all the parameters in the model for the city of Wuhan based on outbreak data published daily by WHO and other sources. In particular they considered the reported confirmed cases in Wuhan from January 23 to February 10 in 2020. As mentioned in their work, they do not consider the change in the way of registering the number of cases occurring from February 12, 2020. Using this approach the authors obtained $\mathcal{R}_0 = 4.25$ as well as the infection risk from each of the three transmission routes, which are, respectively: $\mathcal{R}_1 = 1.959$, $\mathcal{R}_2 = 0.789$, and $\mathcal{R}_3 = 1.497$, indicating that the highest risk of transmission is from exposed (asymptomatic) to susceptible, followed by environment to susceptible and finally from infectious to susceptible, possibly due to isolation measures. They also showed that all the solution orbits converge to the endemic equilibrium, illustrating its global asymptotic stability. That is, the coronavirus infection would remain endemic.

Maier and Brockmann [42] modified the SIR mode to include a new compartment, which accounts for general public containment efforts or individual behavioral changes in response to the epidemic. This transforms the model into the SIR-X which accounts for both quarantine of symptomatic infected individuals and population-wide isolation practices. The model is used to stress the importance of containment strategies to minimize the propagation of the epidemic. Another model that takes into account quarantine was developed by Zhao and Chen [43] and includes compartments for Susceptible, Un-quarantined infected, Quarantined infected, and Confirmed infected. the model is then applied to the analysis of COVID-19 in (i) Wuhan, (ii) Hubei (excluding Wuhan), (iii) China (excluding Hubei) and (iv) four first-tier cities of China. Here again the emphasis has been on proposing rigorous quarantine and control measures at some early times to avoid uncontrolled propagation of the virus.

A more complex model, which starts from the basis of a SEIR scheme, was developed by Godio et al. [44] to study the evolution of the COVID-19 epidemic in Italy. It considers as before a Quarantine compartment to which Infections individuals are directed to. Another interesting characteristic is that the rates at which quarantined individuals are recovered or die are dependent on time. The reason for that assumption is that "the health system can improve its capability to treat people over time, e.g., with the introduction of a new therapy". In Fig. 2.9 we illustrate the scheme of the "generalized" SEIR model used by Godio et al. [44] where we included the rates of recovery and death using the symbols from the paper. The rest of rates are excluded from the graphic to avoid confusion with previous symbols used here. Here the compartment of "Insusceptible" refers to the "part of the population who for various reasons become insusceptible to the disease".

The authors used a stochastic approach in fitting the model parameters using a Particle Swarm Optimization (PSO) solver. Their goal was to improve the reliability of predictions in the medium term (30 days). The time-dependent recovery and death rates where considered as exponential functions based on the data collected by Peng et al. [45]. The assumption here, previously considered by Cheynet [46], is that the death rate should become closer to zero as time increases, while the recovery rate converges toward a constant value. Using their model Godio et al. [44] modeled the evolution of the epidemic in Italy as well as on particular Italian regions starting on 1th March until mid May. They used both the deterministic solution of the "generalized" SEIR model as well the stochastic solution using PSO. They concluded that the deterministic approach is not appropriated to explore the possible solutions of the space-domain due to the underdetermination of the mathematical problem. In contrast, the use of PSO shows some advantages for estimating different scenarios for a 30-day epidemic evolution.

A few models have been developed with a large number of compartments. Wan et al. [47] described the evolution of COVID-19 in mainland China, excluding Hubei province. Their model consists of the following compartments: Susceptible (S), Exposed (E), Infectious with symptoms (I), Infectious but asymptomatic (A), Isolate susceptible isolates (Si), Quarantine infectious pending confirmation (Q), Hospitalized (H) and Recovered (R). This model results in systems of 11 equations interconnecting these compartments. The model has to be fed by the following empirical parameters: (1) Probability of transmission per contact, β; (2) Initial contact rate, c_0; (3) Transition rate of exposed individuals to the infected class, ϕ; (4) Probability of having symptoms among infected individuals, θ; (5) Transition rate of quarantined infected individuals

to hospital class, η; (6) Recovery rate of symptomatic infectious individuals, γ_I; (7) Recovery rate of asymptomatic infected individuals, γ_A; (8) Recovery rate of quarantined infected individuals, γ_H; (9) Disease-induced death rate, d; (10) Correction factor of transmission probability with asymptomatic infectious individuals, ξ; (11) Rate at which the quarantined uninfected contacts were released, μ; (12) Intervention coefficient with respect to contact, q_1; (13) Intervention parameter with respect to patient detection, q_2; (14) Intervention parameter with respect to close contact tracing, q_3; and (15) Exponential decreasing rate of contact rate δ.

The authors obtained an expression for the effective reproductive number based on the total number of cumulative reported cases which is set to be $T(t)$:

$$\mathcal{R}_0(t) = \frac{\beta\theta c_0 e^{-\delta T(t)}}{\gamma_I + d + q_2} + \frac{\beta\xi(1-\theta)c_0 e^{-\delta T(t)}}{\gamma_A}. \tag{2.14}$$

Then, using the estimated parameters they obtained $\mathcal{R}_e(1) \approx 3.34$ on January 20th and then $\mathcal{R}_e(12) \approx 0.89$ on January 31st, 2020, indicating a control of the epidemic in mainland China except Hubei province by the end of January. One of the most interesting conclusions of this study is an alert about what should happen if the quarantine measures were suddenly relaxed before the pandemic was totally controlled. They considered a total population of 1,336,210,000 inhabitants and using their model they simulated a complete removal of quarantine measure by March 5th and by March 20th. The simulations were performed for two different values of the intervention coefficient with respect to contact: $q_1 = 0.2$ and $q_1 = 0.5$ indicating an exponential growth in the last case for both starting dates.

Similarly, Ivorra et al. [48] have considered a model with 8 compartments: Susceptible, Exposed, Infectious that will be detected, Infectious that will not be detected, Hospitalized or in quarantine at home, Hospitalized that will die, Dead, and Recovered. The model is applied to the study of the propagation of COVID-19 in China (including Chinese Mainland, Macao, Hong-Kong and Taiwan). Ivorra et al. [48] estimated the need of beds in hospitals for intensive care units and considered the effects of incomplete data of the results of the model. In addition they estimated the error produced by the model when identifying the parameters at early stages of the epidemic.

In order to account for the complexities of the transmission mechanism of SARS CoV-2 which include the transmission from an infection source, possibly bats, to hosts, currently unknown, to a reservoir, like the Huanan Seafood Wholesale Market, to humans, Chen et al. [49] have constructed a model with 14 compartments. The complexities of this model can be understood by considering that bats and hosts are divided into four compartments each: susceptible (bats, S_B or hosts S_H), exposed (bats, E_B or hosts E_H), infected (bats, I_B or hosts I_H), and removed (bats, R_B or hosts R_H). People were divided into five compartments: susceptible people (S_P), exposed people (E_P), symptomatic infected people (I_P), asymptomatic infected people (A_P), and removed people (R_P) including recovered and death people. The reservoir forms a unique compartment W. The model was further simplified by ignoring the transmission network of Bats-Host and by assuming that the initial value of W is a given impulse function. Using this simplified 6-equations model the authors found that the value of $\mathcal{R}_0$ was 2.30 from reservoir to person and 3.58 from person to person. According to these results the transmissibility of SARS-CoV-2 "was higher than the Middle East respiratory syndrome in the Middle East countries, similar to severe acute respiratory syndrome, but lower than MERS in the Republic of Korea".

Another extension of SEIR is the model developed by Eikenberry et al. [50] which includes susceptible (S), exposed (E), symptomatic infectious (I), hospitalized (H), asymptomatic infectious (A), recovered (R) and death (D) compartments. The most interesting aspect of this model is that the original 7 equations controlling the inter-compartment dynamics are split into 14 equations due to the assumption that some fraction of the general population wears masks. Then, the susceptible population is split into those with S_M and without S_U masks, respectively. The same happens for the individuals in each of the other compartments. The masks are considered to have uniform inward efficiency ε_i, which accounts for the primary protection against catching disease, and outward efficiency ε_o, which accounts for source control/protection against transmitting disease. Then, the basic reproductive number has the following expression:

$$R_0 = \beta\left[S_U + S_M(1-\varepsilon_i)(1-\varepsilon_o)\right]\left(\frac{\alpha}{\varphi+\gamma_I} + \frac{\eta(1-\alpha)}{\gamma_A}\right), \tag{2.15}$$

where:

β: infectious rate;
S_U: proportion of susceptible population without masks;
S_M: proportion of susceptible population with masks;
ε_i: inward efficiency of masks;
ε_o: outward efficiency of masks;
α: fraction of asymptomatic cases;
φ: rate at which asymptomatic are hospitalized;
γ_I : recovery rate of symptomatic infectious individuals;
γ_A : recovery rate of asymptomatic infectious individuals;
η : infectiousness of asymptomatic individuals relative to symptomatic ones.

This work provides a remarkable justification for the use of masks during this pandemic as they found that face masks are useful with respect to both preventing illness in healthy persons and preventing asymptomatic transmission. The authors considered hypothetical mask adoption scenarios, for Washington and New York state, which clearly suggests

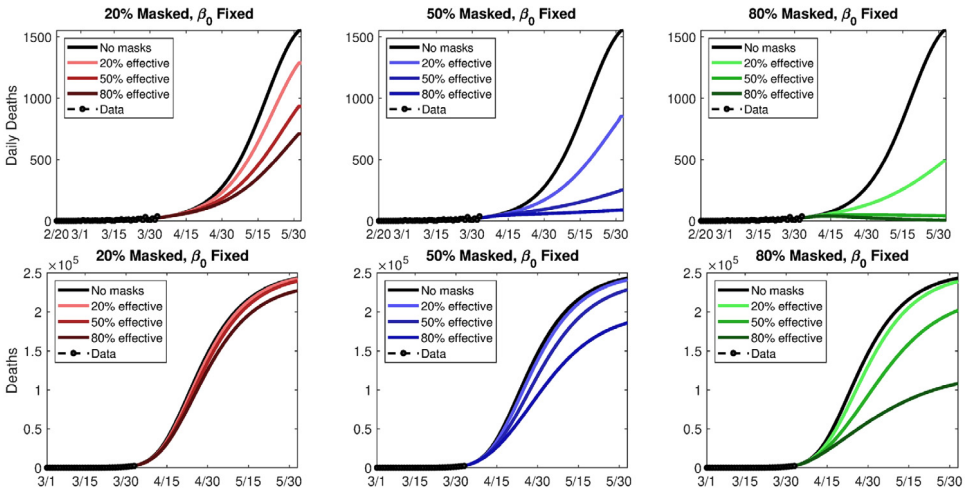


Fig. 2.10. Results of the simulations reported by Eikenberry et al. [50] for the cumulative death tolls for Washington state (top panels) and New York (bottom panels), using a fixed transmission rate, and different permutations of general public mask coverage and effectiveness. Reproduced with permission.

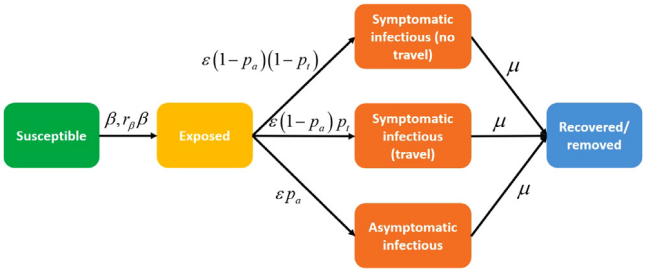


Fig. 2.11. Scheme of the compartments used in the model developed in [53–55].

that the moderate (50% of population using it) or universal (80% of population using it) adoption of masks would have prevented between 17 and 45% of projected deaths over two months in New York. Very interesting, even the use of very weak masks, which have only 20% efficiency, can be useful in reducing the transmission rate as can be seen in Fig. 2.10 when β_0 is fixed.

López and Rodó [51] developed an extended stochastic SEIR model, which takes into account susceptible, exposed, infected, quarantined, recovered, death and confined population. They incorporated time-decaying effects for the parameters controlling the rate at which quarantined population recovered or die, due to potential loss of acquired immunity, people's increasing awareness of social distancing and the use of non-pharmaceutical interventions. They then applied their model to the effects of lockdown and the way in which it is removed to prevent epidemic growth, as well as a potentially larger second wave of SARS-CoV-2 cases occurring within months. Another model consisting of 8 compartments was developed and studied by Giordano et al. [52]. The compartments considered are: susceptible (S), infected (I), diagnosed (D), ailing (A), recognized (R), threatened (T), healed (H) and extinct (E), and the model was collectively termed SIDARTHE. They applied SIDARTHE to the analysis of the epidemic in Italy and compared the results with real data, where they modeled possible scenarios of implementation of countermeasures. Accordingly, they concluded that restrictive social-distancing measures need to be combined with widespread testing and contact tracing to end the ongoing COVID-19 pandemic.

In a series of previous works, the group of Vespignani ([53–55]) developed a compartment model which takes into account several important characteristics of propagation processes like the one of SARS CoV-2. In this model a susceptible individual in contact with a symptomatic or asymptomatic infectious person contracts the infection at rate β, or $r_\beta\beta$, respectively. Then, it enters the exposed compartment where she is infected but is not yet infectious. At the end of the latency period $\varepsilon - 1$, each latent individual becomes infectious, entering the symptomatic compartments with probability $1 - p_a$ or becoming asymptomatic with probability p_a. The symptomatic cases are further divided between those who are allowed to travel (with probability p_t) and those who are not allowed traveling. The model scheme is presented in Fig. 2.11.

Adapting this model to the situation created by the outbreak of COVID-19, Chinazzi et al. [56] show that, at the start of the travel ban from Wuhan on 23 January 2020, most Chinese cities had already received many infected travelers.

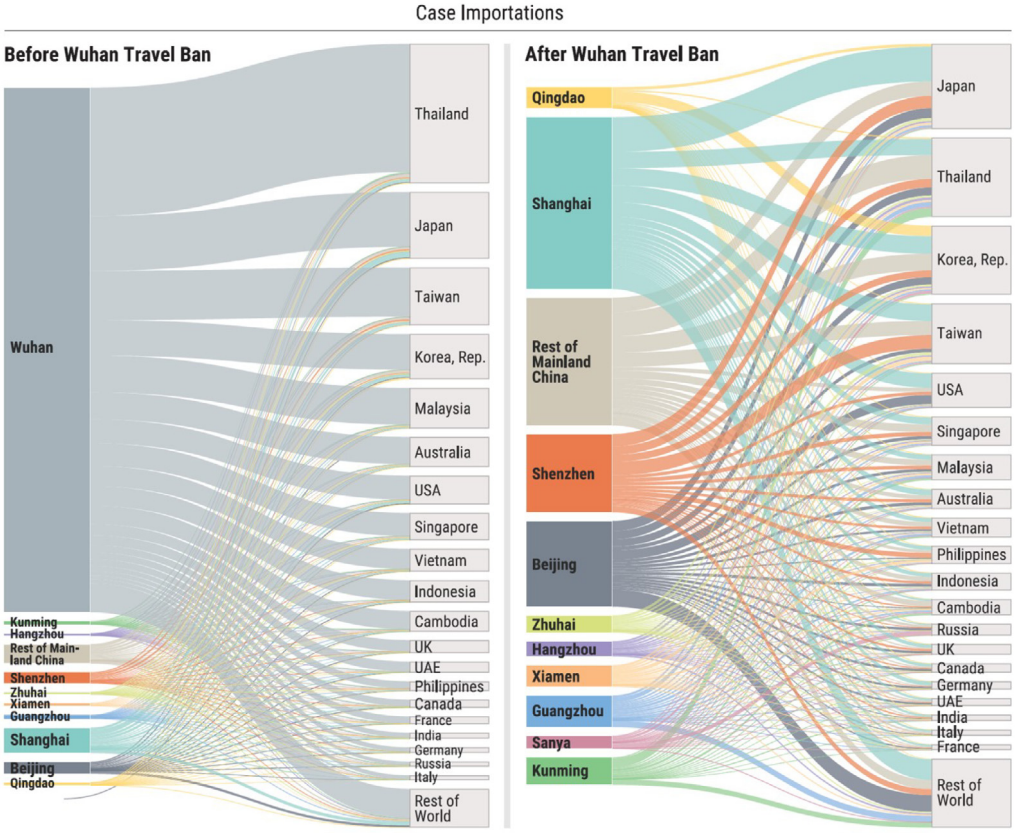


Fig. 2.12. Contribution of cases from the 10 Chinese cities with the highest rates of disease to the relative risk of importation in different countries before (left) and after (right) travel ban in Wuhan as reported by Chinazzi et al. [56].

Consequently, the travel ban from Wuhan delayed only 3–5 days the epidemic progression in mainland China and moved the focus of the epidemic spread outside China from Wuhan to other major cities in the country. Indeed, they have shown that before the travel ban, about 86% of the internationally imported cases originated from Wuhan. However, after the travel ban, the top contributors were Shanghai (28.1%), Beijing (14%), and Shenzhen (12.8%), which accounted for at least 80% of the internationally imported cases (see Fig. 2.12). Their modeling results also indicated that sustained 90% travel restrictions to and from mainland China only modestly affect the epidemic trajectory unless they are combined with a 50% or higher reduction of transmission in the community.

Another compartment model which accounted for human mobility was developed by Li et al. [57] which divided the population in susceptible, exposed, documented infected, and undocumented infected from a total population in a city. They focused on the propagation of COVID-19 within China for which they considered mobility data, a networked dynamic metapopulation model and Bayesian inference, to infer critical epidemiological characteristics associated with SARS-CoV2. The mobility data captured individual movement among the 375 cities simulated in the metapopulation model, using human mobility data from the Tencent location-based service (LBS) used in popular Tencent mobile phone applications (Apps), such as Wechat, QQ, and Maps. They then estimated that 86% of all infections were undocumented prior to 23 January 2020 travel restrictions. Per person, the transmission rate of undocumented infections was 55% of documented infections. They were the infection source for 79% of documented cases. This was one of the first clear alarms about the importance of tracing undocumented cases by massive tests as it was one of the main causes for the rapid geographic spread of SARS-CoV2.

A multi-compartment model which also accounts for different layers in which a population can have social contacts was elaborated by Aleta et al. [58]. The model is built on the basis of a synthetic population divided into three layers (see Fig. 2.13a). The total population considered was of about 85,000 nodes, of which 64,000 are adults and 21,000 correspond to children (see Fig. 2.13b). Before any containment was introduced the model considered about 5M interactions. The analysis is focused on the Boston area, USA, from which data about social contacts was obtained. The multi-compartments scheme is illustrated in Fig. 2.13c, where the population was split into the following compartments: Susceptible (S), Latent asymptomatic (L_A), Latent symptomatic (L_S), Pre-symptomatic (P_S), Infectious asymptomatic (I_A), Infectious symptomatic (I_S), Hospitalized (H), Hospitalized in intensive care (ICU) and Recovered (R) individuals. Aleta et al. [58] considered the

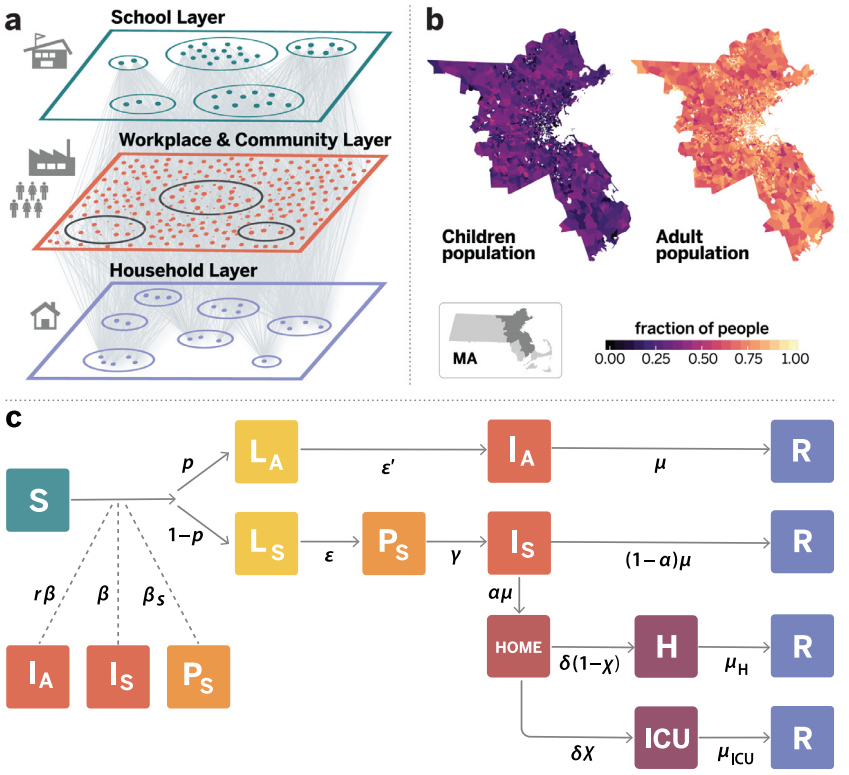


Fig. 2.13. Modeling scheme of work of Aleta et al. [58] where the weighted multilayer synthetic population is built from mobility data in the metropolitan area of Boston (a). The agent-based system of adults and children, whose geographical distributions (b). The compartmental model used (c) where the description of the variables is given in the text. Figure provided by the authors.

effects of three different levels of confinement: (i) school closures; (ii) partial "stay at home; (iii) full lock-down and confinement. They concluded that to avoid a saturation of the health care system it is necessary to implement strict social distancing combined with robust levels of testing, contact-tracing and household quarantine. Interestingly, they found that by identifying 50% of the symptomatic infections, and tracing 40% of their contacts and households, it is assured a reduction in transmission that allows the reopening of economic activities without collapsing the health care system.

Finally, we consider a model that accounts for several important characteristics of the contagion process relevant to COVID-19. The model is developed by Arenas et al. ([59,60]) as an extension of previous metapopulation schemes incorporating recurrent mobility patterns previously developed by the same authors [61–63]. In addition to the consideration of several compartments as in the majority of models, this one also considers an explicit and realistic network of contacts, and a differentiation of age groups during the course of the epidemic. Then, the model is built by dividing the population into the following compartments: susceptible (S), exposed (E), asymptomatic infectious (A), infected (I), hospitalized to Intensive Care Unit, ICU (H), dead (D), and recovered (R). In each compartment the population is divided into N_G age strata: young people (Y), with age up to 25; adults (M), with age between 26 and 65; and elderly people (O), with age larger than 65. Finally, the population is assigned to different geographic regions or patches. A sketch of the model is represented in Fig. 2.14.

Let $\rho_i^{m,g}(t)$ be fraction of agents in the compartment $m \in \{S, E, A, I, H, D, R\}$, which are in the age stratum $g \in N_G$ and which is in the geographic patch i at time t. The system of equations for this model is given below

$$
\begin{aligned}
\rho_i^{S,g}(t+1) &= \rho_i^{S,g}(t)(1-\Pi_i^g(t)),\\
\rho_i^{E,g}(t+1) &= \rho_i^{S,g}(t)\Pi_i^g(t) + (1-\eta^g)\rho_i^{E,g}(t)\\
\rho_i^{A,g}(t+1) &= \eta^g\rho_i^{E,g}(t) + (1-\alpha^g)\rho_i^{A,g}(t)\\
\rho_i^{I,g}(t+1) &= \alpha^g\rho_i^{A,g}(t) + (1-\mu^g)\rho_i^{I,g}(t)\\
\rho_i^{H,g}(t+1) &= \mu^g\gamma^g\rho_i^{I,g}(t) + \omega^g(1-\psi^g)\rho_i^{H,g}(t) + (1-\omega^g)(1-\chi^g)\rho_i^{H,g}(t)\\
\rho_i^{D,g}(t+1) &= \omega^g\psi^g\rho_i^{H,g}(t) + \rho_i^{D,g}(t)\\
\rho_i^{R,g}(t+1) &= \mu^g(1-\gamma^g)\rho_i^{I,g}(t) + (1-\omega^g)\chi^g\rho_i^{H,g}(t) + \rho_i^{R,g}(t),
\end{aligned}
\tag{2.16}
$$

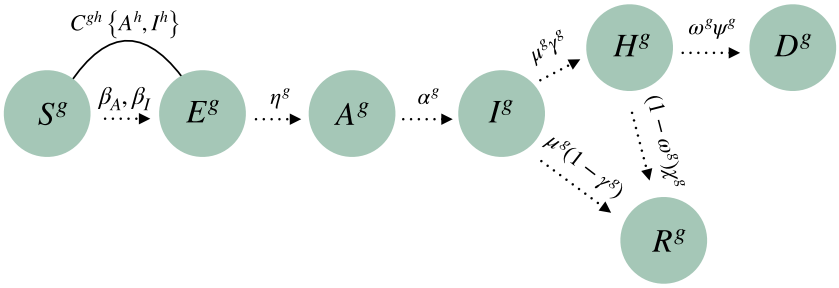


Fig. 2.14. Scheme of the compartment model used by Arenas et al. [59,60]. Figure provided by the authors.

where

Π_i^g: probability of getting infected from asymptomatic or infected individuals;
η^g: probability of exposed individuals of becoming asymptomatic;
α^g: probability of asymptomatic to become infected;
μ^g: escape rate from infectious state;
γ^g: probability of requiring hospitalization at ICUs;
ω^g: probability of individuals in ICUs to die;
ψ^g: rate at which individuals at ICUs die;
χ^g: rate at which ICUs discharge.

The authors then applied their model to the analysis of the epidemic impact in Spain, by considering the autonomous regions in which the country is divided and the mobility between communities, which represents the network connecting the metapopulations of each of the patches or regions. The data about the population and mobility were taken from the National Institute of Statistics of Spain. According to their results the peak of incidence will happen in the first half of April 2020 in absence of mobility restrictions for all Spanish communities (see Fig. 2.15). Accordingly, the results point out to critical situation of the Spanish health capacity system, in particular that for intensive care units, from the end of March. They show different epidemic containment scenarios and conclude that total lockdown was a necessary measure to avoid a massive collapse of the Spanish national health system.

In a subsequent study [59], the authors refined the compartmental model shown above to account for the effects of confinement and social distancing. There, they put the focus on the evolution of the effective reproductive number $\mathcal{R}(t)$ for each subpopulation and age group, allowing to monitor the effects of the different lockdown measures. The integration of these containment measures into the Markovian model yields an effective reproduction number for each municipality and age group, that after some simplifying assumptions reads:

$$\mathcal{R}^g(t) = (\beta_A\tau_A + \beta_I\tau_I)\,(1-\kappa_0(1-\phi))\left(\kappa_0 k^g_{home} + (1-\delta)(1-\kappa_0)k^g_{home+work}\right) \times \sum_{h=1}^{G} C^{gh}\langle\rho^{S,h}(t_c)\rangle\,, \tag{2.17}$$

where κ_0 is the fraction of individuals confined in their households, ϕ is the permeability of the confinement, and δ accounts for the reduction of social contacts for the non-confined population, i.e. the social distancing measures. The effective reproduction number allows to pinpoint the degree of confinement needed to bend the epidemic curve (see Fig. 2.16).

In Table 1 we resume the results previously described in this Review, which use compartmental models for investigating COVID-19 epidemiological parameters.

2.3. Non-compartmental models

Although most of the works reported on modeling the epidemic evolution of COVID-19 are based on compartment models, other approaches, mainly statistical ones, have been used. In one of these works, Wells et al. [64] used maximum likelihood estimation to analyze the impact of international travels and border control on the global spread of COVID-19. They calibrated the daily probability that an infected person would travel outside of mainland China by fitting the predictions of exported cases to reported international incidence for cases that had a travel history to China. The model considers the average duration of the incubation period, and longest time window over which a symptomatic case could travel, which was estimated from an empirical distribution of the duration between symptom onset and first medical visit. Then, the risk of an infected case being exported from Wuhan city was evaluated. The model was also fed from the flight data connecting different cities, such that the probability of travel for an infected person can be estimated. From this, the authors calculated the country-specific risk of importation of an infected individual, which was then validated for 21 countries. Wells et al. [64] also calculated likelihood distributions of the time between arrival and symptom onset to

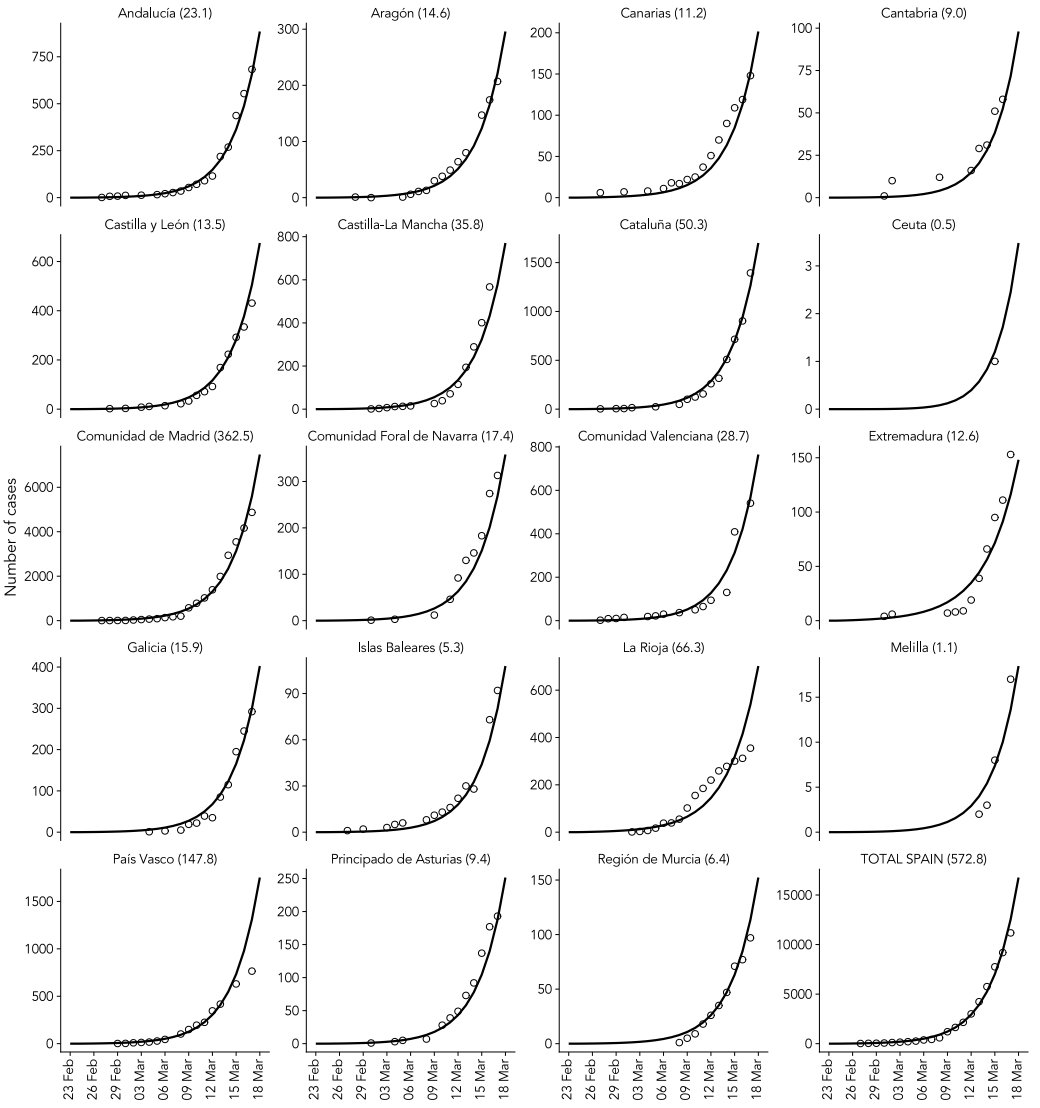


Fig. 2.15. Illustration of the results reported by Arenas et al. [60] for each autonomous region in Spain. The solid line is the result of the epidemic model, aggregated by ages, for the number of individuals inside compartments (H+R+D) that corresponds to the expected number of cases, and dots correspond to real cases reported. Figure provided by the authors.

estimate the incubation period of cases imported. Finally, they also evaluated the effects of taking some border controls to avoid the entrance of infected individuals to a given airport. All in all, this work reports that the risk of exporting cases from mainland China before January 13th 2020 was of 95%, which means 779 cases exported by February 15, 2020 before any border or travel restrictions were imposed. The implementation of such restrictions by Chinese government prevented 70.5% of cases.

A work reported by Wang et al. [65] used statistical and network analysis of 1212 patients in China and infers some epidemiological parameters of the spread of COVID-19 in that country. They started by estimating the incubation period using maximum likelihood estimation from the assumption that the incubation time τ follows the logarithmic normal distribution: $\ln \tau \sim N(\mu, \sigma^2)$, where μ and σ^2 are estimated from data. They estimated average, mode and median incubation time periods are 7.4, 4 and 7 days, respectively, with incubation periods of 92% of people of no more than 14 days. Wang et al. [65] constructed 'transfer' networks where the nodes represent dates between January 5th and February 12th. Two nodes i and j are connected if an individual exposed to the disease at date i developed symptoms or is diagnosed at date j. A weight is assigned to each of the edges indicating the number of patients from a total of 483, which were transferred from one state to another at the corresponding dates. Self-loops corresponds to cases were exposition and symptoms/diagnostic coincide in time. They then calculated the in- and out-degrees for each date, which gives the total number of patients with clinical symptoms or diagnosis, and the total number of exposed persons that will be confirmed

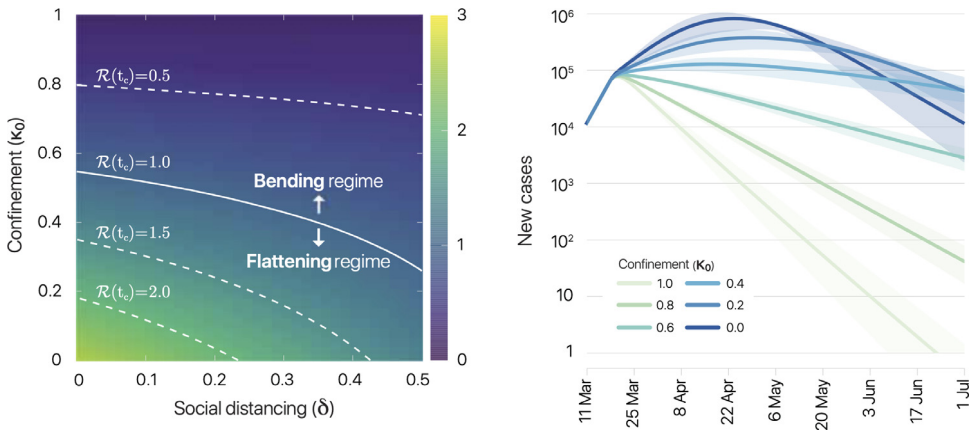


Fig. 2.16. Left: Value of the effective reproduction number, $\mathcal{R}(t_c)$, when containment measures are taken as a function of the confinement κ_0 and social distancing δ computed from Eq. (2.17). Right: Evolution of new cases for different values of confinement κ_0 when social distancing is fixed to $\delta = 0.4$.

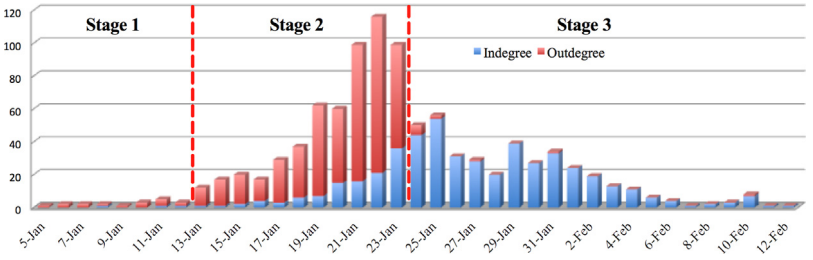


Fig. 2.17. Illustration of the weighted in- and outdegree distributions of the directed graph studied by [65]. Figure provided by the authors. *Source:* Reproduced with permission.

to be infected later, respectively. Interestingly, as illustrated in Fig. 2.17 the plot of the in- and out-degrees of each date reveal some characteristics of the epidemic outbreak. Namely, it revealed the existence of three stages of the epidemic. One stage was before January 13th, where the numbers of both exposed and confirmed patients were low, and the number of exposed (outdegree) patients was higher than confirmed patients (indegree); a second stage, between January 13th and 23rd, where the disease was in outbreak period, with the number of exposed and confirmed cases was high, although the number of exposed cases was higher than the diagnosed ones. The last stage, after January 23rd, was marked by a number of confirmed cases that were higher than that of exposed ones, and the confirmed cases roughly decreased with time. This may be attributed to the adoption of prevention and control measures. Finally, the authors created a network from 1105 patients that have been treated in 248 hospitals, 123 interhospital transfer relationships that involved 206 patients, and 208 patients that were clustering infected. This network contains hospitals and patients as nodes, and patient–patient, hospital–hospital and patient–hospital edges. The first kind of edges represents ties between patients (family friendship, colleagues), the second represents transfers of patients between hospitals and the last represents that the patient was treated in that hospital. This complex system would be much better represented as a multilayer network. From the analysis of this heterogeneous network they concluded that a few hospitals encompass a large number of patients in treatments, and that the aggregate outbreak phenomena were ubiquitous.

Another statistical analysis of the epidemic was performed by Zhao et al. [66] by using correlational analysis. The work focuses on quantifying the association between the load of domestic passengers departed from Wuhan and the number of confirmed cases in a 10 city-clusters. The correlational analysis is based on the model:

$$\log\left[\mathbb{E}\left(c_{i,t}\right)\right] = \alpha_i \cdot \text{province}_i + \beta \cdot \xi_{i,t-\tau} \cdot \varepsilon_{t-\tau}, \tag{2.18}$$

where $c_{i,t}$ is the daily number of new cases in the ith provincial region at time t, $\mathbb{E}(\cdot)$ is the expectation, province_i is a dummy variable for the corresponding province, α_i is a locality-varying interception term, $\xi_{i,t-\tau}$ is the daily number of passengers from Wuhan to the province i, τ denotes the delay from being infected to be detected, and $\varepsilon_{t-\tau}$ is the number of new cases in Wuhan. Using goodness of fit based on the McFadden's pseudo-R-squared, they found $\tau = 5$ days, which agrees very well with the mean incubation period of the infection at 5.2 days. They found a statistically significant positive association between the load of passengers multiplied by the local infectivity in Wuhan and the number of cases reported outside Wuhan. According to their model, increasing by 100 both the number of infected people in Wuhan and the number

Table 1
Resume of the compartmental models used to analyze COVID-19 epidemiological parameters described in the current review.

Authors	Model	Findings	Ref.
D'Arienzo & Coniglio	SIR	$2.43 \leq R_0 \leq 3.10$ for Italy	[17]
Wangping et al.	SIR extended SIR	$R_0 = 3.10$ for Italy, $R_0 = 2.48$ for Hunan, China $R_0 = 4.34$ for Italy, $R_0 = 3.16$ for Hunan, China	[18]
You et al.	SIR	effects of control measures on R_0 for several cities in China	[19]
Roda et al.	SIR/SEIR	found that increasing complexity of the models drops model reliability	[20]
Fanelli & Piazza	SIRD	prediction of peak and number of dead in Italy	[21]
Anastassopoulou et al.	SIRD	estimation of the number of casualties in Hubei, China	[27]
Prasse et al.	SIR + networks	evolution of epidemic with disaggregated β	[23]
Zhou et al. Fang et al.	SEIR	analytic expression of $R_e(t)$	[25,26]
Hou et al.	SEIR	effects of quarantine and isolation on epidemic peak	[27]
Fox et al.	SEIR + hospitalized and ICU cases	$R_0 = 2.4$ without interventions, $R_0 = 1.6$ with nonpharmaceutical interventions in South Wales, Australia	[28]
Grant	SEIR	overestimation of epidemic duration	[30]
Small & Cavanagh	SEIR + networks	necessity of including contact networks	[31]
Linka et al.	SEIR + transport networks	good predictive capacity at earlier stage of the epidemic	[32]
Chung & Chew	SEIR + multiplex	evolution of $R_e(t)$ in Singapore	[33]
Tian et al.	SEIR + statistical inference	effects of lockdown in Wuhan, China	[34]
Manchein et al.	SEIR + data analysis	power-law of epidemic growth across countries	[35]
Peirlinck et al.	SEIR + networks	reports latent and contact periods in China and USA	[36]
Kucharski et al.	SEIR + data analysis	$R_0 = 2.35$ without interventions, $R_0 = 1.05$ one week after contention in Wuhan, China	[37]
Block et al.	SEIR + networks	effects of reducing social contacts on epidemic progression	[38]
Keeling et al.	SEIR + contact tracing	effects of contact tracing on detectability of the disease	[39]
Aleta and Moreno	SEIR + metapopulation + data analysis	epidemic progression at region-level in Spain	[40]
Yang & Wang	SEIRV	$R_0 = 4.25$ for Wuhan, China, disaggregated reproduction numbers for E–I: $R_1 = 1.959$, S–I: $R_2 = 0.789$, and V–I: $R_3 = 1.497$	[41]
Maier & Brockmann	SIR-X	effects of quarantine and asymptomatic on epidemic progression	[42]
Zhao & Chen	SIR + quarantined and unquarantined confirmed infected	effects of rigorous quarantine on epidemic progression	[43]
Godio et al.	SEIR + quarantine and insusceptible	evolution of epidemic in Italy	[44]
Wan et al.	SEIR extended to 11 compartments	$R_e = 3.34$ on January 20th and $R_e = 0.89$ on January 31th in Wuhan, China	[47]
Ivorra et al.	SEIR extended to 8 compartments	necessity of beds in hospitals for ICUs in China	[48]
Chen et al.	SEIR extended to 14 compartments	$R_0 = 2.30$ from reservoir to person, $R_0 = 3.58$ from person to person	[49]
Eikenberry et al.	SEIR extended to 7 compartments	effects on R_0 of wearing face masks	[50]
López & Rodó	stochastic extended SEIR	effects lockdown and time-decaying parameters on disease progression	[51]
Giordano et al.	extended SIR (SIDARTHE)	effects of social measures on the evolution of epidemic in Italy	[52]
Chinazzi et al.	extended SEIR + traveling network	effects of traveling from China on the progress of the epidemic on the world	[56]
Li et al.	SEIR extended to documented and undocumented + mobility	effects of undocumented infections on disease transmission	[57]
Aleta et al.	extended SEIR + multilayer	effects of different contention measures on epidemic in Boston, USA	[58]
Arenas et al.	SEIR extended to 7 compartments + networks + age strata	effects of epidemic on health care system at region-level in Spain and analytical analysis of R_e	[59,60]

of passengers departing from that city, will produce a likely increase of 16.25% in the daily number of cases offsite detected on average.

A very important aspect to consider when modeling disease spreading is that population level parameters, like the basic reproductive number R_0, can hide the relevance of individual variation in infectiousness. This is particularly important

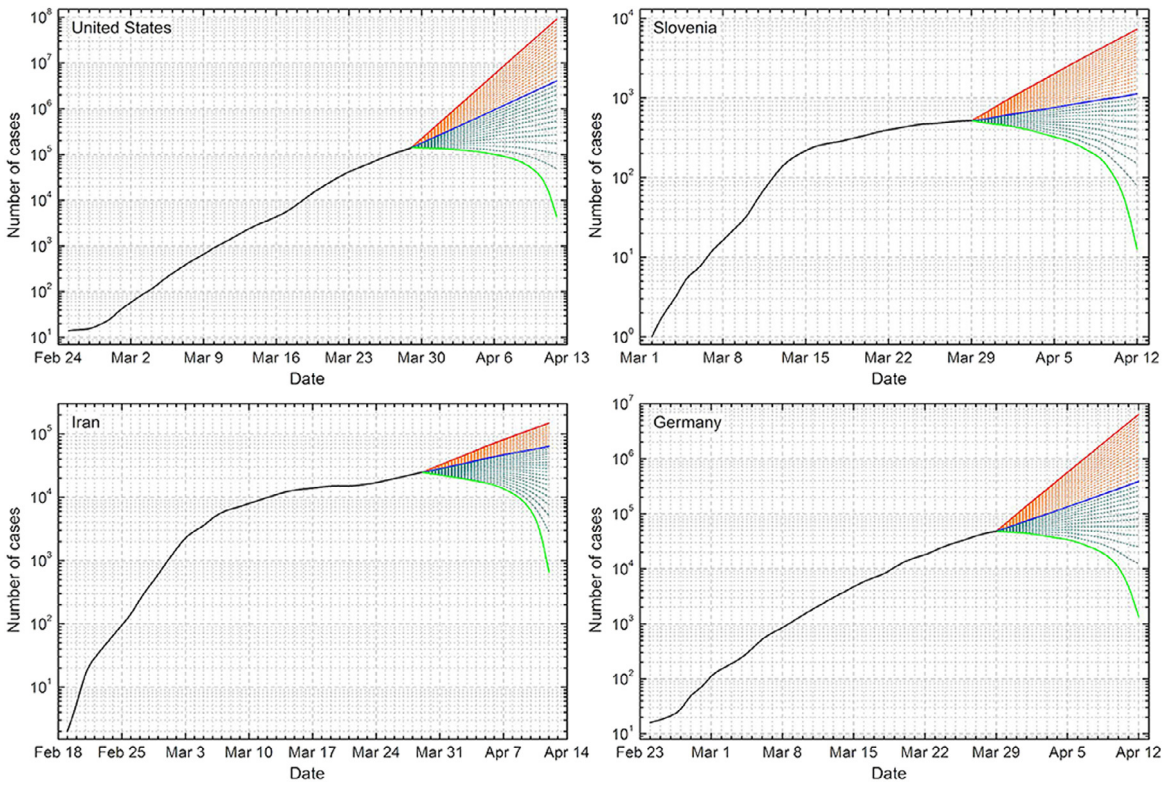


Fig. 2.18. Forecasts made by Perc et al. [69] of COVID-19 cases for the United States, Slovenia, Iran, and Germany. Black solid line denotes the actual data, which were for this analysis last updated on March 29th. The different output scenarios are displayed in different colors: nothing would change (solid blue line); maximal daily growth rate increased by 20% (solid red line); daily growth rate would drop to zero (green line); equally spaced decreasing daily growth rates from top to bottom (orange and olive dashed lines). Figure reproduced with permission of the authors.

when 'superspreading events', where individuals infect unusually large numbers of secondary cases. The investigation of this phenomenon was conducted by Lloyd-Smith et al. [67], who considered branching process and statistical analysis to analyze the influence of individual variation in infectiousness on disease emergence for several diseases. They used contact tracing data from eight directly transmitted diseases, and showed that the distribution of individual infectiousness around R_0 is often highly skewed. According to their modeling results individual-specific control measures outperform population-wide measures. Another study based on branching process model was developed by Hellewell et al. [68] where the number of potential secondary cases produced by each individual was obtained from a negative binomial distribution with a mean equal to the reproduction number. The model also uses heterogeneity in the number of new infections produced by each individual. This study focuses on assessing the feasibility of contact tracing and case isolation to control outbreaks of COVID-19. The authors simulated new outbreaks starting from 5, 20, or 40 introduced cases, and concluded that contact tracing and isolation might not contain outbreaks of COVID-19 unless very high levels of contact tracing are achieved. More importantly, if there is a high fraction of transmission from asymptomatic infected individuals, the model shows that this strategy might not achieve control within 3 months.

In a work published by Perc et al. [69], the authors developed a simple iteration method based on the daily values of confirmed cases as the only input to forecast the progression of COVID-19 in different countries. They determined maximally allowed daily growth rates to explore different scenarios which includes departing from the exponential growth. Accordingly, the authors show that daily growth rates should be kept at least below 5% to converge to plateaus in short times as illustrated in Fig. 2.18 for different countries.

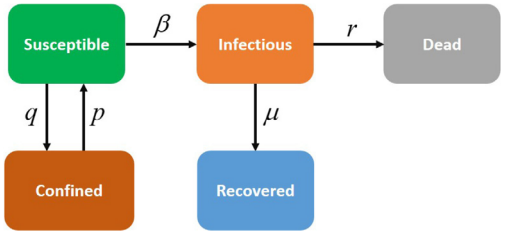
Finally, we briefly describe a model developed by researchers at the Imperial College of London [29], which is based on previous developments by [70]. The model has reached wide diffusion through academic[3] and nonacademic media. It is based on stochastic, spatially structured individual-based simulations, in which at any time-step of $\Delta T = 0.25$ days, a susceptible individual i has probability $1 - \exp(-\lambda_i \Delta T)$ of being infected, where λ_i is the instantaneous infection risk for individual i, which depends on several parameters. Individuals can be infected in household, places, or by random contacts in the community. The model is fed by household size and age structure data, school size data and school allocation model,

[3] See for instance: https://www.nature.com/articles/d41586-020-01003-6.

workplace data, and commuting distance data. This group has been applying this model to respond to different questions related to the COVID-19 and posting systematic reports about their findings.[4] By July 2nd, there were 29 Reports published by the Imperial College group. The model sometimes gives very wide margins for the parameters estimated, such as for the number of infected individuals in Spain and Italy, where the predictions in Report 13 by [71] were: Spain 15% [3.7%-41%] and Italy 9.8% [3.2%-26%] for the percentage of total population infected (mean [95% credible interval]) (see Discussion for the comparison with more recent results).

2.4. On the predictability of epidemiological models

An important question related to the epidemiological models previously considered is whether they can accurately predict some important future events of the epidemic evolution. In particular, Castro et al. [72] addressed the question of predictability focusing on two important events: the turning point and the end of an expanding epidemic. They concluded that these events cannot be accurately predicted because "exponentially growing dynamics are intrinsically unpredictable" in a similar way as chaotic events. They considered a variation of a SIR model in which a new class C of individuals sent to confinement is considered. These individuals are susceptible but not infected. They also consider a compartment for people who die, which is different from those who recover from the disease. The general scheme and equations of the model are presented below:



$$\begin{aligned}
\dot{S}(t) &= -\beta S(t) I(t) - pS(t) + qC(t), \\
\dot{C}(t) &= pS(t) - qC(t), \\
\dot{I}(t) &= \beta S(t) I(t) - (r+\mu) I(t), \\
\dot{R}(t) &= rI(t), \\
\dot{D}(t) &= \mu I(t).
\end{aligned} \tag{2.19}$$

Then, using data for the epidemic in Spain the authors applied a Bayesian approach to fit the data, assuming that the numbers of infected, recovered and dead are log-normally distributed with unknown variance and mean given by the expression for $I(t)$ obtained from the previous model. Their results are illustrated in Fig. 2.19. As can be seen percentiles 5% and 95% provide contradictory results: either the epidemic curve dies out or it will keep growing exponentially, albeit at a different rate. The authors remarked that this is a consequence of the inherent variability of the fitted parameters as summarized by the posterior distributions and the exponential character of the epidemic.

Castro et al. [72] also analyzed the distribution of the day in which the model predicts the maximum peak of the epidemic in Spain conditioned on it actually occurring. Accordingly, it is concluded that the model can infer neither the peaking time nor whether there is a peak at all. From further analysis of the predictability of this class of models the authors concluded that the problem is not produced by the quality of data, although better data is certainly needed, but that the problems arise because "small variations in the parameters bring about growing uncertainties as time elapses". Therefore, the problem is not solved by adding more variables or compartments as the problem of the susceptibility to initial condition will persist.

As we have resumed before, the statistical approaches, which are intrinsically probabilistic, fully rely on past data to predict the near future. As stressed by Castro et al. [72] these approaches "only yield likelihoods of different scenarios, with intervals of confidence that grow extremely fast as time elapses". They consider that the model, previously described here, elaborated by researchers at Imperial College, "yields the most trustable (probabilistic) predictions to date" because they are compatible with multiple scenarios in most countries in the mid-term. However, the huge intervals of confidence inherent to their approach also limit predictability to the near future.

[4] The Reports can be reached at: https://www.imperial.ac.uk/mrc-global-infectious-disease-analysis/covid-19/covid-19-reports/.

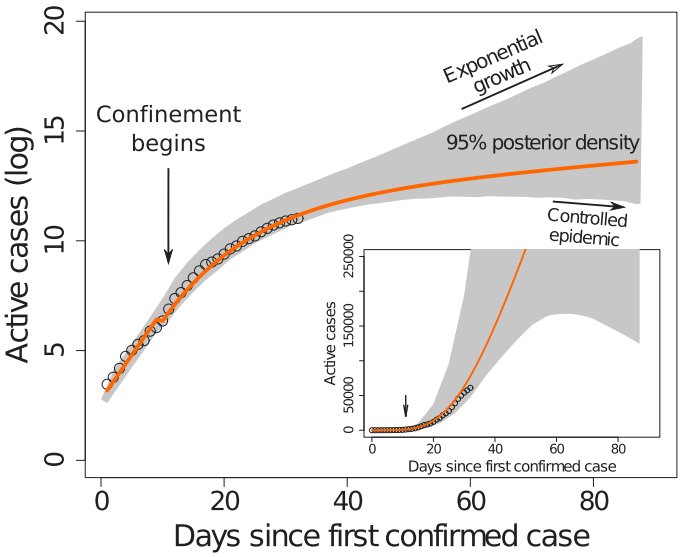


Fig. 2.19. Fit to data for the daily number of active cases in Spain from March 1st to March 29th as reported by Castro et al. [72]. The solid line represents estimation using the median parameters for each posterior in the model. The shaded area represents the 95% predictive posterior interval. Inset: same data and curves with linear vertical scale. The figure is reproduced with permission of the authors.

3. Modeling for drug repurposing

Currently, there are no specific drugs against SARS CoV-2, the virus that produces COVID-19. In the period between 2012 and 2017, the Food and Drug Administration (FDA) of USA approved a total of 24 new antiviral agents, i.e., in the categories of new molecular entity or new combination [73]. From these antiviral drugs 11 were for treating HIV-1, 10 for treating Hepatitis C virus, one for treating influenza, one for Hepatitis B virus, and one for Cytomegalovirus. In total, in the ten years between 1987 and 2017, FDA approved 108 antiviral drugs, none of them is for treatment of a coronavirus. These numbers do not parallel the research done in these years about the development of new potent antiviral compounds. This situation is not alien to the whole modern pharmaceutical research which is characterized by a high attribution rate, where only 10% of compounds that arrive at Phase II clinical trials are eventually approved as a drug. Most of the failures are due to safety concerns or poor efficiency.

An alternative route for drug discovery which has increased its popularity in recent years is drug repurposing (also known as drug repositioning, reprofiling or re-tasking). It consists in strategies for identifying new uses for previously approved or investigational drugs [74–78]. That is, a new use outside the scope of the original medical indication is found for a previously developed drug. The advantages of this approach are that the repurposed drug: (i) has previously being found to be sufficiently safe in humans, (ii) most of the preclinical tests, safety assessments, formulation development, pharmacokinetics, administration–distribution–metabolism–excretion (ADME), etc. of the repurposed drug are known or in advanced stages of development. Therefore, the time and investment needed for repurposing a known drug to a new use are significantly smaller than those for developing a new one.

In the case of antiviral drugs the following scenarios of drug repurposing have been identified [79]:

(1) Same target–new virus: Consists in finding activity against other virus for a known antiviral drug for which is known its specific viral target or cellular function/pathway. For instance, favipiravir which was approved for treating influenza and repurposed against Ebola and Zika viruses;
(2) Same target–new indication: Consists in identifying a non-antiviral drug that acts on a pharmacological target, e.g., a protein or a pathway, which exists in a virus, therefore exploiting this drug as an antiviral therapeutic agent. For instance, the anticancer drug imatinib that is under studies to be repurposed against pathogenic coronaviruses;
(3) New target–new indication: Consists in finding a non-antiviral drug with a specific target not existing in viruses to have a new molecular target which exists in a virus. For example, the antimicrobial ivermectin which shows inhibition of viral replication.

Apart from imatibib, which was previously mentioned, in 2018 there were other two drugs under study to be repurposed against MERS- and SARS-CoV. They are GS-5734 (an investigational antiviral drug) under studies in nonhuman primates, and chlorpromazine (approved antipsychotic) under study in infected cells in vitro.

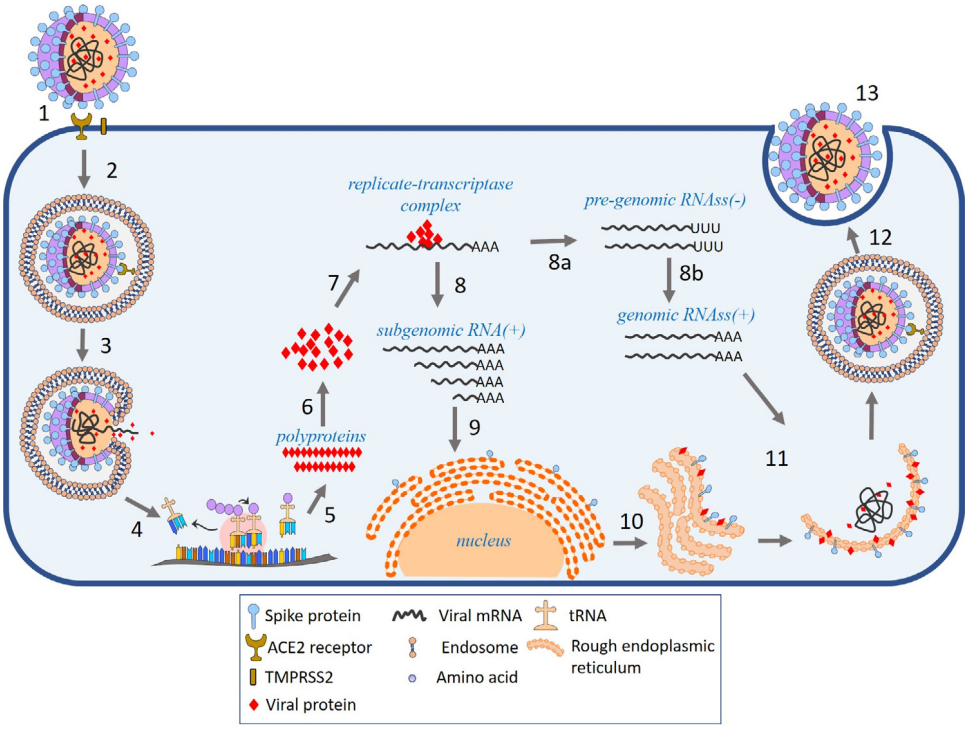


Fig. 3.1. Illustration of the life cycle of the SARS CoV-2 based on current knowledge. The graphics were prepared using Motifolio (https://www.motifolio.com/).

In general, drug repurposing consists of the following three steps:

- hypothesis generation, where the candidate molecule is identified;
- mechanistic assessment of the drug effects in preclinical models;
- evaluation of efficacy in phase II clinical trials.

The hypothesis generation is a critical step of the whole process and several computational techniques have been developed to make this process more efficient. Some of these computational approaches are described below, but before we resume some of the most important facts about the biology of SARS CoV-2 which are essential for drug repurposing.

3.1. SARS CoV-2

The virus that produces COVID-19 is a *Betacoronavirus* known as SARS CoV-2. It is an enveloped, single-stranded positive-sense RNA genome virus of a spherical shape with a diameter between 60 and 140 nm. The life cycle of SARS CoV-2 is illustrated in Fig. 3.1 (see [80–83]). It starts when SARS CoV-2 enters human cells by means of receptor-mediated endocytosis mechanism, in which its spike glycoprotein S, previously preprocessed by the transmembrane serine protease 2 (TMPRSS2), binds to the host receptor Angiotensin-Converting Enzyme-2 (ACE2) (1). In the interior of the human cell, the virus enters in the form of an endosome (2), where cathepsin L, an endosomal acid protease, activates the spike protein by cleaving the protein into S1 and S2. At this point, the S2 protein fuses the membrane of the virus with that of the endosome (3), which results in the release of the viral material, RNA and proteins, into the human cell cytoplasm. Now, viral mRNA is translated (4) by using the human cell ribosome to form two viral replicase polyproteins: pp1a and pp1ab (5). These two polyproteins are then cleaved by the Papain-like protease and 3C-like (main) protease, resulting in 16 non-structural proteins (nsp) (6). A replicase–transcriptase complex is then assembled with some of these proteins and the viral RNA (7). It drives the production of subgenomic RNA(+) through transcription (8) and of pre-genomic RNAss(-) (8a) which is replicated to genomic RNAss(+) (8b). The subgenomic RNA(+) is now translated into structural proteins (9), such as the spike S protein, envelope proteins (E), membrane proteins (M), nucleocapsid proteins (N), and several Open reading frame proteins (Orf). All of them enter the endoplasmic reticulum (10) where the nucleoprotein complex is formed between the nucleocapsid protein and the (+) strand genomic RNA. Finally, the assembly of all proteins and RNA into a new virus particle is carried out in the Golgi apparatus of the human cell (11). The virus is then expelled from the cell via exocytosis (12) and starts its maturation to start a cycle again (13).

The 29 proteins produced by SARS CoV-2 are described in Table 2 (see [84]). They constitute potential pharmacological targets for drug repurposing, which is a process described in the next subsection.

Table 2
Proteins in the SARS CoV-2, their names and brief description of their function.

No.	Protein	Description
1	nsp1	Suppresses host antiviral response
2	nsp2	
3	nsp3	nsp3–nsp4–nsp6 complex involved in viral replication
4	nsp4	nsp3–nsp4–nsp6 complex involved in viral replication
5	nsp5	Main protease (3C-like)
6	nsp6	nsp3–nsp4–nsp6 complex involved in viral replication
7	nsp7	nsp7–nsp8 complex is part of RNA polymerase
8	nsp8	nsp7–nsp8 complex is part of RNA polymerase
9	nsp9	ssRNA binding
10	nsp10	Essential for nsp16 methyltransferase activity
11	nsp11	Short peptide
12	nsp12	RNA polymerase
13	nsp13	Helicase/triphosphatase
14	nsp14	3'–5' exonuclease
15	nsp15	Uridine-specific endoribonuclease
16	nsp16	RNA-cap methyltransferase
17	S	Spike protein, mediates binding to ACE2
18	Orf3a	Activates the NLRP3 inflammasome
19	Orf3b	
20	E	Envelope protein, involved in virus morphogenesis and assembly
21	M	Membrane glycoprotein, predominant component of the envelope
22	Orf6	Type I IFN antagonist
23	Orf7a	Virus-induced apoptosis
24	Orf7b	
25	Orf8	
26	N	Nucleocapsid phosphoprotein, binds to RNA genome
27	Orf9b	Type I IFN antagonist
28	Orf9c	
29	Orf10	

3.2. Repurposing based on molecular docking

Molecular docking consists on predicting, by means of structure-based computational strategies, the interactions between a ligand and a target. Some potential targets have been identified in the literature for SARS CoV-2 [85]. The target is typically a receptor in the virus, e.g., a protein. The first stage for the drug repurposing process is the identification of potential receptor targets in the virus. Once the receptor is known, it is necessary to identify its potential binding sites, then multiple drugs could be interrogated against that specific binding site in the target. According to De Clercq and Li [86], 3-chymotrypsin-like (main) protease, papain-like protease, helicase, RNA-dependent RNA polymerase and the spike protein are attractive targets to develop antiviral agents against SARS CoV-2.

We should distinguish here between two different scenarios. In the first scenario we consider the existence of the three-dimensional (3D) structure of the protein target in a complex with an inhibitor. In this case we already can directly apply docking algorithms for a series of known drugs, such as those in large libraries of ligands, into the binding site where the inhibitors have been found experimentally. Such 3D structures can be obtained either by Nuclear Magnetic Resonance (NMR), X-rays crystallography or cryo-microscopy and are commonly deposited in the Protein Data Bank (PDB) [87]. In the second scenario we consider that such 3D structure of potential targets does not exist and they should be built by comparative models. Once, such structures are obtained it is necessary to find the potential binding sites in which inhibitors can be docked. We will consider both scenarios here for the case of SARS CoV-2.

By the time of writing this article (24th June, 2020) the structures of 5 out of the 29 proteins of SARS CoV-2 have been reported forming complexes with inhibitors. Notice that other structures of SARS CoV-2 proteins have also been solved but not in the presence of inhibitors (see next Subsection). They are (see Fig. 3.2) the main or 3CL protease (a) [88], papain-like protease (b) [89], nsp15 endoribonuclease (c) [90], nsp12–nsp7–nsp8 complex bound to the template-primer RNA (d) [91] and nsp16–nsp10 heterodimer (e) (to be published).

3.2.1. Inhibitors of the main protease

The 3D structure of the main protease of SARS CoV-2 was resolved using X-rays crystallography by several authors. Some of these structures are for complexes between Mpro and inhibitors that have been found to bind the active site of the main protease. In this case it is possible to directly apply docking strategies for drug repurposing. The main goal here is to predict the best conformations/orientations of a ligand within the protein binding site. Docking process consists in effectively sampling the conformational space described by the free energy landscape to find conformations/orientations that minimize a given scoring function, which should associate the native bound-conformation to the global minimum of the energy hypersurface. Liu and Wang [92] have classified the existing scoring functions into four categories: (i) "force-field-based" or "physics-based", (ii) "empirical" or "regression-based", (iii) "knowledge-based potential", and

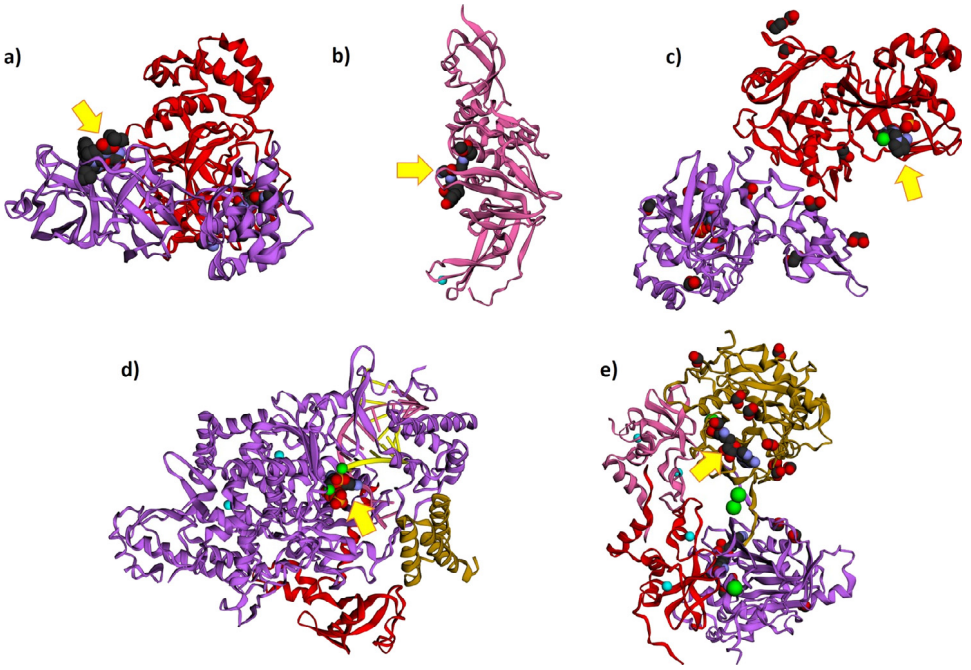


Fig. 3.2. Crystal structures of 3CL protease (a), papain-like protease (b), nsp15 endoribonuclease (c), nsp12–nsp7–nsp8 complex bound to the template-primer RNA (d) and nsp16-nsp10 heterodimer (e) with inhibitors. The inhibitors are marked with a yellow arrow. The PDB structures correspond, respectively, to: 6Y2G [88], 6WX4 [89], 6WXC [90], 7BV2 [91] and 6WKQ (to be published).

(iv) "descriptor-based" or machine-learning based" ones. In the studies reported so far for drug-repurposing against SARS CoV-2, the scoring functions that have been used belong to the categories (i) and (ii). A "physics-based" scoring function has the following general form:

$$\triangle G_{\text{binding}} = \triangle E_{\text{vdw}} + \triangle E_{\text{electrostatic}} + \triangle E_{\text{H-bond}} + \triangle E_{\text{desolvation}}, \tag{3.1}$$

where the $\triangle E$ terms account for the energy change due to: van der Waals, electrostatic, hydrogen bond and desolvation interactions between the ligand and the protein. Empirical scoring functions compute the fitness of protein–ligand binding as the sum of contributions of different individual terms which represent some important energy factor of the binding.

In a recent work, Jin et al. [93] determined the structure of the main protease Mpro of SARS CoV-2 bounded to the inhibitor denoted as N3. In another study, Zhang et al. [88] determined the 3D structure of the main proteinase of SARS CoV-2 bounded to an α-ketoamide inhibitor. These structures have been the template for several studies of virtual screening against the main protease of SARS CoV-2. In one of these studies, Ton et al. [94] used a modification of the molecular docking protocol Glide,[5] known as Deep Docking, to virtually screening 1.3 billion compounds from ZINC15 library[6] against the main protease of SARS CoV-2. Using this deep learning platform that provides fast prediction of docking scores, they identified 1000 potential ligands for SARS-CoV-2 Mpro. In another study, Fischer et al. [95] also performed docking studies of a large dataset of compounds on Mpro of SARS CoV-2. In this case they screened a library of over 687 million compounds. They combined this search with molecular dynamics simulations used to validate the stability of the ligand–Mpro complexes which resulted in a list of 11 drug-like compounds with improved binding free energy to the target protease relative to the inhibitor N3. The same 3D structure was used as the basis for the virtual screening of drug-like compounds using an advanced deep Q-learning network with a fragment-based drug design strategy [96]. The authors reported a series of 47 lead compounds which can be used as potential candidates for researchers in their development of drugs against SARS-CoV-2.

The previously mentioned studies cannot be considered as examples of drug repurposing as they are lead development studies. However, the previously mentioned study of Jin et al. [93] also included a virtual screening strategy to identify known drugs to be repurposed against SARS CoV-2. They used the N3–Mpro complex as a model for identifying lead inhibitors using in silico screening based on Glide. They interrogated chemical compounds from an in-house database and found that cinanserin fits adequately into the substrate-binding pocket. Cinanserin is a well-known serotonin antagonist discovered in the 1960s and which was previously identified as an inhibitor of SARS CoV. The half-maximum

[5] A docking program that uses an empirical scoring function. See https://www.schrodinger.com/glide.

[6] https://zinc15.docking.org/.

inhibitory concentration of this compound for Mpro is $IC_{50} = 125\mu M$. In another repurposing study Khan et al. [97] used molecular docking to identify potential hits followed by studies of molecular dynamics simulation and binding free energy calculations to evaluate the dynamic behavior, stability of protein–ligand contact, and binding affinities of the hit compounds. They found that three FDA approved antiviral drugs: Remdesivir (broad spectrum antiviral), Saquinavir (anti-HIV), and Darunavir (anti-HIV), are promising hits against SARS-CoV-2 main protease. Finally, we mention the work of Tsuji [98] who interrogated the molecules on the ChEMBL database[7] to find inhibitors of the SARS CoV-2 main protease. In this case the virtual screening analysis was carried out on the basis of the 3D structure of Mpro bounded to an α-ketoamide obtained by Zhang et al. [88]. Then, Tsuji [98] screened all 127,561 distinct compounds in the database using a molecular docking program that uses a "physics-based" scoring function and found 64 potential drugs. These drugs included 11 FDA approved, 14 clinical and 39 preclinical drugs. The approved drugs include the antibacterial: sulfamethizole, sulfathiazole, kanamycin, tobramycin, and phthalylsulfathiazole; the neuropsychiatric: droperidol, eszopiclone and homotropine; the antineoplastic: alpelisib; the cardiovascular: tizanidine; and the gastrointestinal: mannitol.

Hall Jr and Ji [99] used the 3D structure of Mpro with the inhibitor N3 (PDB ID: 6LU7) to perform in silico docking modeling to find drugs to be repurposed as effective inhibitors for SARS-CoV-2. Several antiviral medications: Zanamivir, Indinavir, Saquinavir, and Remdesivir are found as potential hits on the SARS CoV-2 main proteinase. Another study which focused on targeting the main protease was reported by Joshi et al. [100]. They performed screening of 7100 molecules including active ingredients present in the Ayurvedic antitussive medicines, anti-viral phytochemicals and synthetic antivirals against SARS-CoV-2 main protease. Accordingly, they identified several natural molecules like d-viniferin, myricitrin, taiwanhomoflavone A, lactucopicrin 15-oxalate, nympholide A, afzelin, biorobin, hesperidin and phyllaemblicin B that strongly binds to SARS-CoV-2 main protease.

Unveiling the inhibition mechanism of mpro. In an analysis of 92 crystal structures of complexes between the protease of SARS CoV-2 and inhibitors, Nguyen et al. [101] have identified 13 different binding sites in this protein. However, the binding pocket, which is around the catalytic site is the most populated one by inhibitors. That is, about 74% of the complexes between inhibitors and Mpro display the inhibitor located at this binding pocket. There is no other binding site which is occupied by more than 6%, and most of them are occupied by only one of the 92 inhibitors considered. The authors identified some important characteristics of the main binding pocket. For instance, Gly143 residue in Mpro is the most attractive site to form hydrogen bonds, followed by Cys145, Glu166, and His163, all in the main binding pocket. There are 45 out of 92 targeted covalent bonding inhibitors. The analysis of these 92 protein-inhibitor complexes was carried out using a reduced representation protocol based on algebraic topology. This approach produces a dramatic simplification of the geometric complexity of ligand–protein complexes by representing the systems as simplicial complexes. The concepts involved in this approach are the following [102–105]:

Simplex σ_q. A q-simplex is the convex hull of $q+1$ affinely independent points in $\mathbb{R}^n$. The 0, 1, 2, 3-simplex corresponds to single vertex, edge, triangle, and tetrahedron, respectively.

Simplicial complex, K. A set of simplices satisfying that every face of a simplex $\sigma_q \in K$ is also part of complex K, and the nonempty intersection of any two simplices in K is the common face of both.

Persistent homology. The p-persistent qth homology group of K_t is

$$H_q^p(K_t) := Z_q(K_t) / \left(B_q\left(K_{t+p}\right) \cap Z_q(K_t)\right), \tag{3.2}$$

where $\emptyset = K_0 \subseteq K_1 \subseteq K_2 \cdots \subseteq K_m = K$ is a filtration of the simplicial complex K, $Z_q(K)$ is the q-cycle group of K defined by $Z_q(K) = ker\left(\partial_q\right) = \left\{c \in C_q(K) \,\middle|\, \partial_q c = \emptyset\right\}$ and $B_q(K)$ is the q-boundary group of K defined by $B_q(K) = im\left(\partial_{q+1}\right) = \left\{\partial_{q+1} c \,\middle|\, c \in C_{q+1}(K)\right\}$. The rank of $H_q^p(K_t)$ counts the number of q-dimensional holes in K_t that remain in K_{t+p},which is known as the p-persistent qth Betti number (see [106] for a review). An illustration of the process for a protein is provided in Fig. 3.3 taken from Cang and Wei [107].

In another work, Estrada [108] compared the 3D structure of the main protease of SARS CoV-2 with that of SARS CoV-1. In this case the author uses a reduced representation of the protein structures based on graph/network theory, which is known as protein residue networks (PRN). A PRN (see [109] Chapter 14 for details) is built from the information reported on the Protein Data Bank [87]. The nodes of the network represent the α-carbon of the amino acids. Then, a cutoff radius r_C is considered, which represents an upper limit for the separation between two residues in contact. The distance r_{ij} between two residues i and j is measured by taking the distance between C_α atoms of both residues. Then, when the inter-residue distance is equal or less than r_C both residues are considered to be interacting and they are connected in the PRN. The adjacency matrix A of the PRN is then built with elements defined by

$$A_{ij} = \begin{cases} H\left(r_C - r_{ij}\right) & i \neq j, \\ 0 & i = j, \end{cases} \tag{3.3}$$

where $H(x)$ is the Heaviside function. An example for the M^{pro} of SARS CoV-2 is illustrated in Fig. 3.4.

Several 3D structures of the Mpro in their apo form, i.e., no inhibitor present, of SARS CoV-1 and of SARS CoV-2 were transformed into their PRNs. Several graph-theoretic invariants were compared for both kinds of proteases, finding that they differ in less than 2%. A couple of invariant describing the capacity of the protein to transmit perturbations at

[7] https://www.ebi.ac.uk/chembl/.

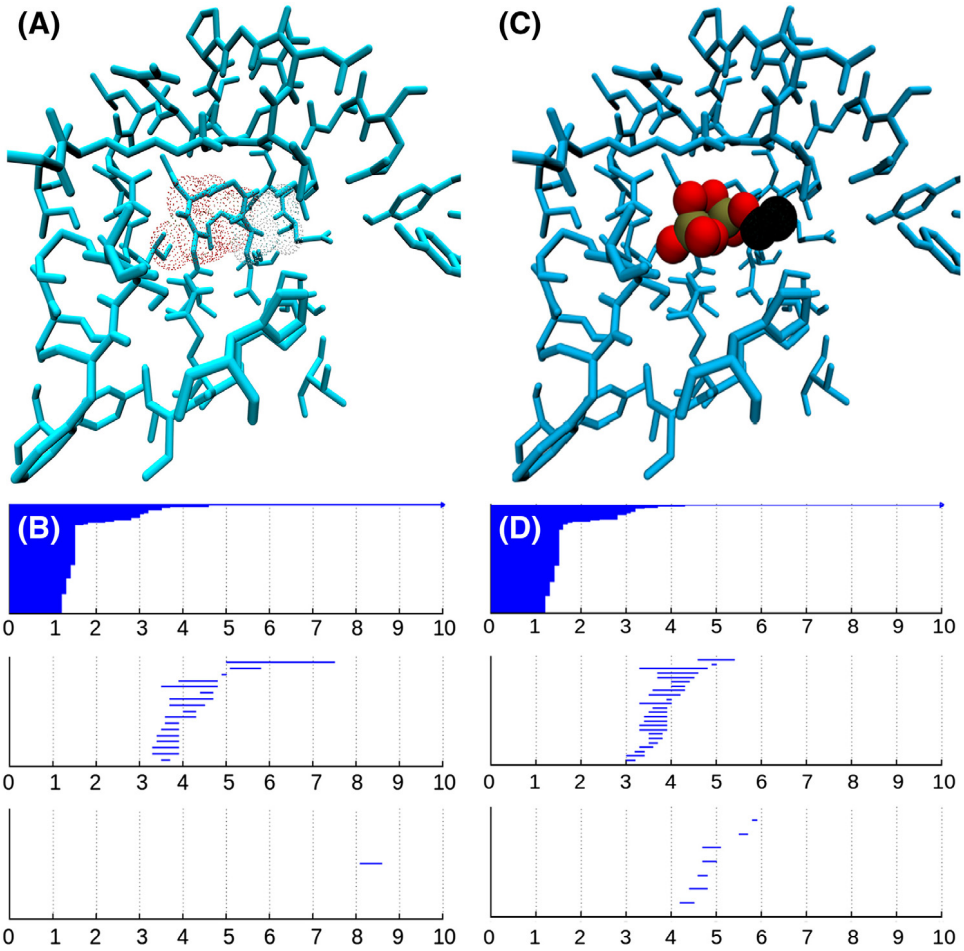


Fig. 3.3. Illustration of protein–ligand binding induced topological fingerprints change as reported by Cang and Wei [107]. Figure provided by the authors.

relatively short distances from the center of the perturbed residue shown differences of 20% between both proteases. These results agree with previous findings that reported that the similarity between the amino acid sequences of both proteases is 96%, and that the superposition of the 3D structures of them displays very little deviations.

However, a remarkable difference between both proteases was found when the transmission of perturbations at long-range distances was considered. In this case, [108] used a graph-theoretic invariant which is able to account for the transmission of perturbation at longer distances through the PRN. These descriptors are mathematically defined on the basis of the following matrix function [110]:

$$Z := \sum_{k=0}^{\infty} \frac{A^k}{k!!} = \frac{1}{2}\left[\sqrt{2\pi}\,\mathrm{erf}\left(\frac{A}{\sqrt{2}}\right) + 2I\right]\exp\left(\frac{A^2}{2}\right), \tag{3.4}$$

where A is the adjacency matrix of the PRN, its kth powers count the number of walks of length k between pairs of (not necessarily different) nodes, and $1/k!!$ penalizes the walks of length k by $k!!$ (double factorial). The main effect of the use of this double factorial penalization instead of the simple factorial.

The SARS CoV-2 proteases are 1900% more sensible to transmit such perturbations than their analogues of SARS CoV-1. This property reflects the improved capacity of the new protease of transmitting perturbations across its domains. The author also found that the amino acids displaying such increased sensitivity to perturbations are around the binding site of the new protease and close to its catalytic site. Estrada [108] also analyzed a few structures of the SARS CoV-2 protease bounded to inhibitors, two of them being the most potent ones reported so far. He found that this increased sensitivity to perturbations observed in SARS CoV-2 Mpro is related to the effects of powerful protease inhibitors. In fact, the strongest inhibitors of the SARS CoV-2 main protease are those that produce the least change of this capacity of transmitting perturbations across the protein. The effects of these three inhibitors on the descriptor accounting for long-range perturbations across the PRN are illustrated in Fig. 3.5 for the close neighborhood around the binding site of

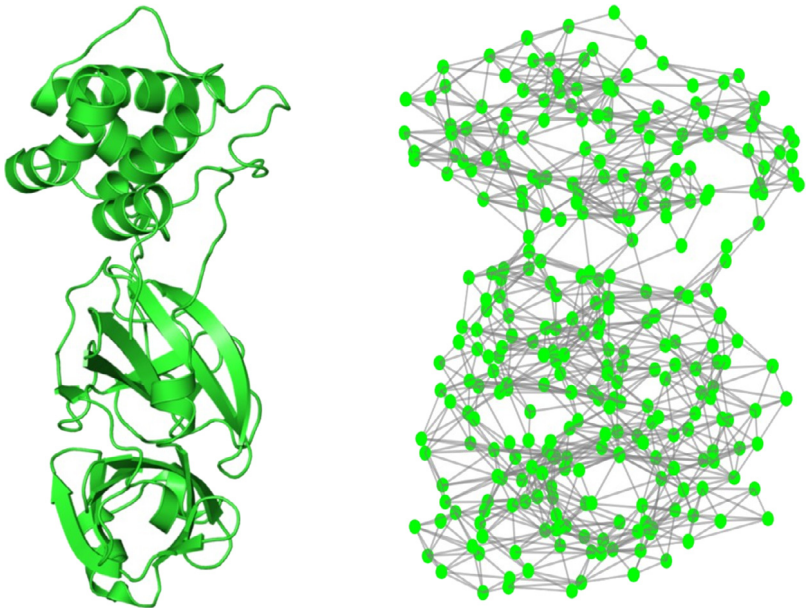


Fig. 3.4. Cartoon representation (left) of the M^{pro} of SARS CoV-2 (PDB ID: 6Y2E) and the corresponding protein residue network (right).

SARS CoV-2 main protease. This strategy developed by Estrada [108] based on PRN and network descriptors was recently applied by Chen et al. [111] to the analysis of seven existing antibodies for SARS-CoV-2 spike (S) protein with three-dimensional (3D) structures deposited in the PDB. Five antibody structures associated with SARS-CoV were evaluated for their potential in neutralizing SARS-CoV-2.

In order to explain the mechanism by which these perturbations are transmitted across the structure of the main protease, Abadias et al. [112] developed a fractional Susceptible–Infected (SI) model based on the assumption that there are similarities between epidemic spreading and a diffusive process on a protein residue network to prove the capability of propagating information in complex 3D protein structures [113]. The new fractional SI model on a network was defined as:

$$D_t^{\alpha}\left(-\log\left(\vec{1}-I\right)\right)(t)=\beta^{\alpha}AI\,(t)\,, \tag{3.5}$$

with initial condition $I\,(0)\,=\,I_0$, and where β is the infection rate, A is the adjacency matrix, $\vec{1}$ is an all-ones column vector, and D_*^{α} is the Caputo fractional derivative defined as

$$D_*^{\alpha}f\,(x)\equiv\frac{1}{\Gamma\,(\tau-\alpha)}\int_0^t(t-s)^{\tau-\alpha-1}\left(\frac{d}{ds}\right)^{\tau}f\,(s)\,ds, \tag{3.6}$$

where $\alpha\,>\,0$, and where $\tau\,=\,\lceil\alpha\rceil$ is the smallest integer greater than or equal to α. The solution of this fractional SI model is expressed, under given initial conditions, by $E_{\alpha,1}\,(\zeta A)$, where ζ is a parameter that depends on the time and on β, and $E_{\alpha,1}\,(\zeta A)$ is the Mittag-Leffler matrix function of ζA,

$$E_{\alpha,1}\,(\zeta A):=\sum_{k=0}^{\infty}\frac{(\zeta A)^k}{\Gamma\,(\alpha k+1)}, \tag{3.7}$$

where $\Gamma\,(\cdot)$ is the gamma function. According to their results, when $\alpha\,=\,0.5$ but not when $\alpha\,=\,1$, some significant and physically sounded trends are observed. First, the most powerful inhibitor increases by 71% the transmissibility of perturbations through the main protease after its binding. It is followed by the second most powerful inhibitor, which increases modestly the transmissibility of perturbations by 13%. However, the weakest inhibitor does not increase, but decreases, the transmissibility of perturbations across the protein. In addition the average length of the shortest paths connecting the pairs of residues with the largest increase in the transmissibility of effects follows the same trend as the inhibitory potency. The most potent inhibitor perturbs an average of 9 residues per perturbation path. The second most powerful inhibitor perturbs an average of 8 residues per shortest paths, and the weakest inhibitor perturbs only 6.

3.2.2. Inhibitors of 2'-O-ribose methyltransferase

Another important target against SARS CoV-2 is 2'-O-ribose methyltransferase (2'-O-MTase), nsp16, which plays an important role in viral replication and prevents recognition by the host innate immune system. The structure of nsp16

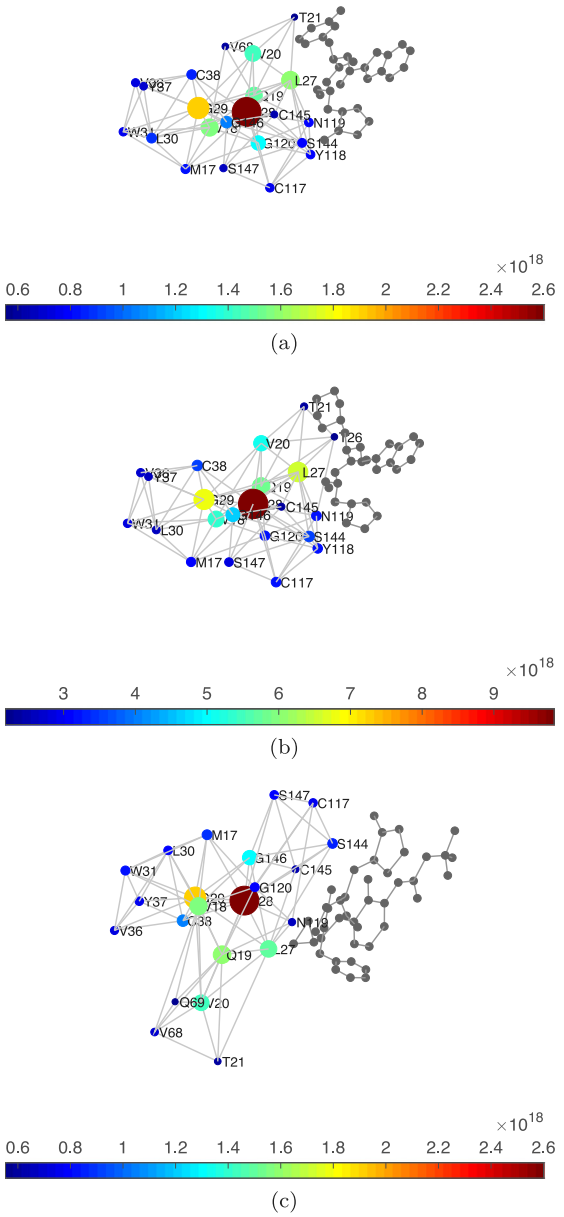


Fig. 3.5. Illustration of the 22 amino acids with the largest values of the long-range subgraph centrality in 6M0K (a), 6LZE (b) and 6Y2G (c). The residues are connected if they are at no more than 7.0 Å. The color bar and the radius of the nodes indicates the values of Z_{ii} normalized to the largest value in the corresponding protein.

in complex with nsp10 and S-adenosyl-L-methionine (SAM) was solved by X-rays crystallography (PDB ID: 6W4H) as illustrated in Fig. 3.6 [114]. SAM is used by nsp16 as the methyl group donor and the catalytic KDKE region (K46, D130, K170, and E203 in SARS-CoV-2) is well conserved. Indeed, previous studies in SARS-CoV-1 demonstrated that the 2'-O-MTase activity is completely removed by single mutation of any residue in this region. Therefore, this binding pocket is a potentially interesting one for inhibiting the activity of this protein and so of the SARS CoV-2.

Jiang et al. [115] analyzed the druggability of all binding sites in the structure of nsp16 using the PDB structure 6W4H described before. According to geometrical and physicochemical properties of potential binding sites in this protein, the SAM pocket was reported to have the highest drug score. Then, they interrogated, using AutoDock Vina,[8] the database Drugslib[9] which contains 7173 stereoisomers corresponding to 4574 'approved' drugs. They found the following

8 http://vina.scripps.edu/.

9 http://www.druglib.com/.

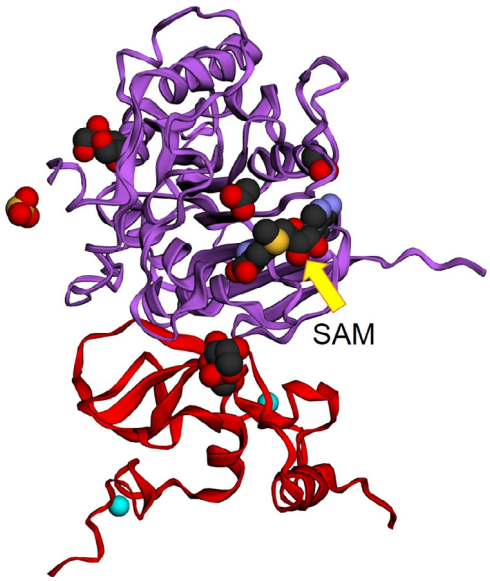


Fig. 3.6. 3D structure of the complex between nsp10 and S-adenosyl-L-methionine (SAM) (PDB ID: 6W4H).

candidates for dug repurposing: MK3207 (CGRP receptor antagonist), Rimegepant (CGRP receptor antagonist), Entrectinib (anti-cancer), Osi-027 (dual inhibitor of mTORC1 and mTORC2), Bolazine (synthetic androgen/anabolic steroid), R428 (potent and selective inhibitor of Axl), hesperidin (flavonone glycoside found in citrus fruits), losulazine (antihypertensive), rebastinib (inhibitor of the TIE2 immunokinase), and Cep-32496 (inhibitor of BRAF). Jiang et al. [115] also explored molecules structurally similar to SAM and found that Gs-9667, trabodenoson, binodenoson, sonedenoson, regadenoson, metrifudil, and selodenoson are potential inhibitors of nsp16.

Other studies [116,117] have used homology modeling to construct the structure of SARS CoV-2 nsp16 protein based on structures previously reported for the same protein in SARS CoV-1. Both studies used virtual screening by molecular docking into the SAM binding site of nsp16. In the first of them Sharma et al. [116]) report 22 drugs that accommodate into the binding site, which includes alkaloids, antivirals, cardiac glycosides, anticancer, steroids and other drugs. The two antiviral agents reported to binding to nsp16 are Saquinavir and Indinavir, which are both inhibitors of HIV protease. In the second study by Khan et al. [117] the authors interrogated an in-house library of 123 antiviral drugs. They also considered the inhibition of the main protease of SARS CoV-2. Then, they reported two known antiviral drugs as candidates for repurposing: Dolutegravir and Bictegravir. Both drugs are known as anti-HIV agents which inhibit the HIV replicase.

3.2.3. Inhibitors of spike protein

The 3D structure of the S protein of SARS CoV-2 has been obtained by using cryo-electron microscopy at 3.5-Å-resolution (PDB ID: 6VSB) [118]. It has been remarked that this trimeric structure may not be appropriate for understanding the receptor binding mechanisms, because some structural information of important parts of the protein are missing [119]. Then, the complete structure of the monomeric and trimeric S protein was constructed by homology modeling using three SARS CoV-1 S proteins as template (PDB ID: 5X5B, 6ACG, 5I08) [119]. These structures were then used by de Oliveira et al. [120] as the basis for a drug repurposing study using molecular docking. They used molecular dynamics simulations of the trimeric structure of the S protein enclosed in a box of 147,803 water molecules. Then, the most stable structures resulting from the molecular simulation experiments were used as the targets for docking calculations using AutoDockTools[10] software. Once the best – according to a scoring function – inhibitors were found, the inhibitor-S-protein complexes were submitted again to molecular dynamics simulations to obtain their optimal conformations. An interesting observation from the molecular dynamics studies is that the receptor-binding domain (RBD) can flip from "up" to "down" conformation into direction of the S-protein center. The last conformation impedes the interaction with the human receptor ACE2, which indicates that a transition "down" to "up" should take place before binding to the surface of human cells. It also shows that RBD domain is very sensitive to the chemical environment, temperature and solvent effects which may affect its conformation. Therefore, in finding drugs that bind this protein it is important to consider such conformational sensibility of the S-protein. The authors carried out a drug repurposing experiment and found, apart from a few traditional herbal isolates, 10 drugs of use in human or veterinary medicine that inhibit the SARS CoV-2 S protein. They found, for instance, that ivermectin may bind in the RBD region of the S

[10] http://autodock.scripps.edu/wiki/AutoDockTools.

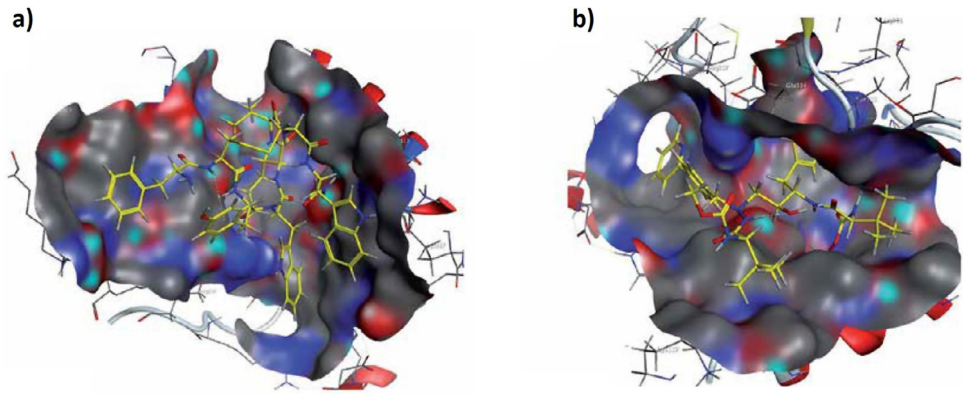


Fig. 3.7. Complexes of vapreotide (a) and atazanavir (b) with SARS CoV-2 helicase protein built from molecular homology reproduced from Borgio et al. [125].

protein, thus inhibiting the coupling of this protein with the human ACE2 receptor. This FDA-approved anti-parasitic has been found experimentally to be an inhibitor of SARS-CoV-2 [121], with a single addition to Vero-hSLAM cells 2 h post infection with SARS-CoV-2. It was able to produce ∼5000-fold reduction in viral RNA at 48 h. The other drugs found by de Oliveira et al. [120] shows a large molecular weight, with an average of 745, such as quinupristin (antibiotic) with MW of 1022 or acetyldigitoxin (cardiac glycoside) with MW 899.

In a different study Wei et al. [122] directly used the 3D structure of SARS CoV-2 protein S determined by cryo-electron microscopy (PDB ID: 6LZG) to detect binding sites using Discovery Studio 2016.[11] They considered as a target the pocket found near the protein–protein interface. They then interrogate molecules in the DrugBank[12] which includes 2628 approved small drugs, which after removal of drugs with molecular weights larger than 500 kDa, including polypeptides, consists of 2191 FDA-approved drugs. They also virtually screened drugs in an in-house database. Among the most interesting findings of their work it is the report that the HIV antiretroviral drug, raltegravir, inhibits the protein S of SARS CoV-2. Another candidate to drug repurposing identified was digotoxin, a cardiac glycoside, which is analogous of acetyldigitoxin identified in the study of de Oliveira et al. [120], and somehow confirming the observation of these last authors that because the S-protein is a glycoprotein it "has high affinity with oligosaccharides and various sugar chains".

Another approach was used by Villoutreix et al. [123] who focused on the furin cleavage site (see [124]) at the S1/S2 boundary of the S protein of SARS CoV-2. Interestingly, this site appears neither in SARS CoV-1 nor in other SARS-related CoV, and it is believed [119] to be responsible for the efficient spread of SARS-CoV-2. They performed molecular docking of 8,000 molecules acting in different therapeutic areas into the furin X-rays structure co-crystallized with a peptide-like inhibitor (PDB ID: 5XH) or co-crystallized with a small chemical compound (PDB ID: 5MIM). They found that the drug sulconazole, which is a broad-spectrum antifungal agent, binds to the active site of furin catalytic domain. Then, the authors found experimentally that sulconazole inhibits the cleavage of the cell surface furin substrate MT1-MMP, which contains two furin cleavage sites similar to those of the SARS-CoV-2 spike protein. This makes sulconazole an interesting candidate for drug repurposing against SARS CoV-2.

3.2.4. Inhibitors of the nsp13 helicase or nsp12 polymerase

Helicase is a viral replication enzyme in coronavirus, which has the characteristics of unwinding DNA and RNA and separating them into two single-stranded nucleic acids. Borgio et al. [125] consider nsp13 helicase as a potential target for drug repurposing. They followed the classical scheme in which starting from the helicase amino acid sequence the 3D structure of the protein is built by homology. Then, after validating the 3D structure, they retrieved drugs from DrugBank and PubChem database[13] and performed molecular docking and energy computation for the complexes inhibitor-nsp13. They found that the analgesic, which is used in treating AIDS-related diarrhea, vapreotide was the best candidate for binding into the nsp13 helicase. It is followed by atazanavir (HIV protease inhibitor), and hydroxychloroquine (antimalarial), but also accompanied by a few antiviral drugs, such as lopinavir, nelfinavir, saquinavir, indinavir, ritonavir, among others. Of the 26 best fitted drugs, 19 are in use against HIV, which points out directly to a strategy for drug repurposing of anti-HIV drugs against SARS CoV-2 helicase. Notice that several of these drugs bind into different binding pockets of the helicase, which gives more choices for the design of molecular dissimilar compounds (see Fig. 3.7).

On the other hand, the polymerase nsp12 forms a complex with nsp7 and nsp8 which is essential for viral replication and transcription (see [126]), and it is regarded as a potential target to fight against SARS CoV-2. The structure of this

11 https://www.3dsbiovia.com/products/collaborative-science/biovia-discovery-studio/.

12 https://www.drugbank.ca/.

13 https://pubchem.ncbi.nlm.nih.gov/.

complex with the inhibitor drug remdesivir is illustrated in Fig. 3.2(d). However, before the structure of this complex was published in the PDB, Ruan et al. [127] constructed it by homology modeling based on the fact that the amino acid sequence alignment of nsp12 of SARS-CoV-2 and that of SARS CoV-1 share 96.35% similarity. Then, they used the structure of the complex in SARS CoV-1 (PDB ID: 6NUR) as a template. They performed molecular docking at the nsp12–nsp7 and nsp12–nsp8 interfaces, respectively. During docking, nsp7 (or nsp8) was removed from the complex and only nsp12 was left as the receptor. They interrogated 7496 drugs against both potential receptors. The authors found seven drugs: Saquinavir, Tipranavir, Lonafarnib, Tegobuvir, Olysio, Filibuvir and Cepharanthine as potential hits to inhibit the polymerase of SARS CoV-2. All these compounds are antiviral drugs, except the last one which is an anti-inflammatory and antineoplastic compound. Saquinavir, tipranavir, lonafarbib and tegoburir were found to bind at the interface between nsp12 and nsp7, while the other drugs were found at a site between nsp12 and nsp8. Notice that in the crystal structure of remdesivir with the nsp12–nsp8–nsp7 complex (see Fig. 3.2(d)) the drug is not bounded at the interface between two proteins but in the core of nsp12, close to the site in which the protein interacts with RNA. This lack of coincidence is a consequence of the fact that Ruan et al. [127] constrained the search of the binding site only to the intersection between the pairs of proteins. A study revealing the capacity of molecular docking to detect precisely the binding sites of drugs on the nsp12 polymerase was conducted by Mirza and Froeyen [128], who also built its structure by homology with that of SARS CoV-1. This study was not directed to the identification of drugs for repurposing but mainly to lead discovery. What is important from the current discussion is that the authors found that the binding site for these lead compounds include the residues: Arg553, Arg555, Asp618, Asp623, Arg624, Ser682, Ser759, Asp760 an Lys798. In the crystal structure of remdesivir bounded with the complex nsp12–nsp8–nsp7, the drug is bounded to the residues: Arg553, Arg555, Cys622, Asp623, Ser682, Thr687, Asn691, and Asp760, which means that they coincide with the predicted ones.

3.2.5. Inhibitors of other target proteins

The 3D structure of SARS CoV-2 papain-like protease was built by Arya et al. [129] using homology modeling and used to interrogate a total of 2525 FDA approved drugs from the DrugBank database or the Zinc15 library. They found that the binding site containing the S3/S4 pockets is more spacious than the one containing the catalytic residues, and it was used for the docking modeling of the drugs. The S3/S4 pocket is formed by residues Asp164, Val165, Arg166, Glu167, Met 208, Ala246, Pro247, Pro248, Tyr 264, Gly266, Asn267, Tyr 268, Gln269, Cys217, Gly271, Tyr273, Thr301 and Asp302. The authors reported several drugs that inhibit the papain-like protease of SARS CoV-2, such as: biltricide (anthelmintic), cinacalcet (calcimimetic), procainamide (antiarrhythmic), terbinafine (antifungal), pethidine (narcotic analgesic), labetalol (to treat hypertension), tetrahydrozoline (over the counter eye drops and nasal spray), ticlopidine (inhibitor of platelet aggregation), ethoheptazine (opioid analgesic), and levamisole (antihelminthic), among others. They found that chloroquine (antimalarial) shows inhibition of the papain-like protease, but its ligand efficiency was low (zero on a scale from 0 to 2).

In a different strategy to the previously discussed here, in which the authors mainly focus on one protein target of SARS CoV-2, Wu et al. [130] considered a systematic homology modeling of the structures of 19 SARS CoV-2 proteins. Then, they performed virtual ligand screening of compound libraries like the ZINC drug database. They then reported lists of drugs acting on: (i) papain-like proteinase; (ii) 3C-like main protease; (iii) RNA-dependent RNA polymerase; and (iv) helicase. Among the drugs reported we can find: (i) Ribavirin, Valganciclovir, β-Thymidine; (ii) Lymecycline, Chlorhexidine, Alfuzosin; (iii) Valganciclovir, Chlorhexidine, Ceftibuten. Beck et al. [131] used a similar strategy to identify antiviral drugs that can be repurposed against SARS-CoV-2. They constructed the structures of 3C-like proteinase, RNA-dependent RNA polymerase, helicase, 30-to −50 exonuclease, endoRNAse, and 20-O-ribose methyltransferase. In general, they showed that atazanavir (anti-HIV) displayed the best results, with a high inhibitory potency against the SARS-CoV-2 3C-like proteinase. It was followed by remdesivir, efavirenz, ritonavir, and dolutegravir. Interestingly, lopinavir, ritonavir, and darunavir which are designed to target viral proteinases were also found to bind to the replication complex components of SARS-CoV-2.

3.3. Signature matching, genome-wide association, pathways and network mapping

Signature-matching strategies for drug repurposing consist in comparing the unique characteristics or 'signature' – transcriptomic, proteomic or metabolomic data; chemical structures; or adverse event profiles – of a drug candidate with those of known drugs, diseases or clinical phenotypes. One of the strategies used here is the 'guilt-by-association' one [132,133], in which if two diseases share some similar therapies, then other drugs that are currently used for only one of the two may also be therapeutic for the other. This can be implemented in a network-based framework as follows. First, create a bipartite network consisting of drugs, e.g., FDA-approved drugs, and diseases. The drugs and diseases are connected if the first is used for treating the second. Now, we can project this bipartite network into the disease–disease space, creating a weighted graph of diseases, where two diseases are connected by a weighted edge representing the number of drugs used for their treatment. Novel drug uses can then be suggested based on shared treatment profiles from any disease pairs. The highest the weight between two diseases the larger the probability of repurposing drugs used for one of them to treat the other disease [132]. In another approach, repositioning can be proposed on the basis of shared transcriptomic signatures between two drugs. In this case the two drugs may share a therapeutic application independently of their structural similarity/dissimilarity. Another signature matching approach is based on studying the

chemical similarities between drugs and assuming that two chemically similar molecules can share the same biological target. Therefore, this approach captures ligand-based similarities among what would otherwise be considered disparate proteins.

A genome-wide association study aims to identify genetic variants associated with common diseases, such that new insights about the biology of the disease can be obtained which may help in finding new targets from which drugs can be repurposed. In a similar way, large-scale identification of common pathways between different diseases or viruses can be identified [134]. Then, it is possible to identify drugs targets through database and literature searches. Network techniques can be used to build drug or disease networks based on gene expression patterns, disease pathology, protein interactions or genome-wide association studies.

In an outstanding effort, Gordon et al. [84] performed a chemo-proteomic exhaustive analysis of the protein–protein interaction (PPI) network of SARS CoV-2 and humans. These authors identified 332 high-confidence SARS CoV-2-human PPIs. Then, they continue their analysis by searching for ligands that may disrupt SARS CoV-2-human PPI. They considered the ''importance'' of a ligand by (i) the statistical significance of the interaction between the human and viral proteins, (ii) the status of the drug (approved, investigational new drug, preclinical), (iii) drug selectivity, and (iv) drug availability. They used chemoinformatic as well as target- and pathway-specific searches of drugs inhibiting any of the human proteins in the SARS CoV-2-human interactome. These searches yielded: 29 approved drugs, 12 investigational new drugs, and 28 preclinical candidates. In total 63 out of the 332 potential targets have molecules that modulate them as can be seen in Fig. 3.8.

In order to complete the investigational cycle of drug repurposing to find agents against SARS CoV-2, Gordon et al. [84] implemented a medium-throughput experimental screening protocol to test some of the candidates found in the chemo-proteomic analysis. In total they tested 47 out of the 69 compounds previously identified, plus a few others identified by other approaches. According to their experimental analysis, there are two classes of molecules in reducing viral infectivity. They are, the inhibitors of protein biogenesis, such as zotatifin and ternatin-4 (both are selective inhibitors of the eukaryotic translation initiation factor 4A), and those which inhibits Sigma1 and Sigma2 receptors, such as haloperidol (antipsychotic), PB28 (Sigma 1/2 modulator) and hydroxychloroquine (antimalarial).

A different strategy, although based on the same data, was followed by Gysi et al. [135]. They focused on the observation that most approved drugs do not target directly disease proteins (DP), but they bind proteins which are in the vicinity of DP, in a network theoretic sense. Therefore, their goal was to identify drugs that have the potential to 'perturb' the network vicinity of the virus disease module independently on whether they target or not a protein to which the virus binds. They defined 12 different protocols (called pipelines in the paper) to quantify the proximity or similarity between a drug–target and a virus target, either in a PPI network or in embedding spaces resulting from the transformation of the own PPI network. The last set of approaches which uses techniques of Artificial Intelligence (AI) based on graph convolutional networks [136], resulted in the best results in the calibration tests performed by the authors. Such calibration tests were carried out by training the different approaches to detect those 67 drugs already in clinical trials against COVID-19. The results of this combined protocols generate 12 different rankings of drugs according to the network structural criteria encoded in each of them. Consequently, Gysi et al. [135] disentangled these rankings by using a rank aggregation algorithm based on Kemeny consensus. As a result of these experiments the authors arrived at a list of 86 drugs which are candidates for repurposing against SARS CoV-2. The candidate that top ranked in this list is ritonavir, an antiretroviral used against HIV, followed by isoniazid (anti-tuberculosis), troleandomycin (antibiotic), cilostazol (cardiovascular) and chloroquine (anti-malarial).

Another example of a network-based drug repurposing study was published by Zhou et al. [137] before the structure of the PPI network of Human-SARS CoV-2 had been determined experimentally. The authors started from the phylogenetic analyses of 15 human coronaviruses (HCoV) whole genomes and found that SARS-CoV-2 shares the highest nucleotide sequence identity with SARS-CoV-1 (79.7%). Specifically, they found that the envelope and nucleocapsid proteins are two evolutionarily conserved regions. Then, the authors searched for HCoV-host interactome networks and found 119 host proteins associated with CoVs. These host proteins are either the direct target of HCoV proteins or are involved in important pathways of the coronaviruses infection. Using these identified proteins Zhou et al. [137] reported 47 human proteins that can be targeted by at least one approved or experimental drug under clinical trials. The most targettable proteins identified were: GSK3B, DPP4, SMAD3, PARP1, and IKBKB. They then constructed a drug–target network by assembling target information for more than 2000 FDA-approved or experimental drugs and computationally identified 135 drugs that were associated with the HCoV-host interactome. They finally identified 16 repurposable drugs, which include: mesalazine (against inflammatory bowel disease), toremifene (anticancer), eplerenone (steroid), paroxetine (antidepressant), sirolimus (immunosuppressant), dactinomycin (anticancer), irbesartan (cardiovascular agent), mercaptopurine (anticancer), melatonin (sleep hormone), and others. A problem with this work is that the proteins reported here as the most targettable ones, and many others, are not among the ones found experimentally by Gordon et al. [84] as the ones interacting with SARS CoV-2 proteins. Also, those most important proteins found by Gordon et al. [84] are not among the ones studied by Zhou et al. [137]. Therefore, the necessity for reviewing the existing techniques for identifying potential protein targets from computational methods is evident for future studies against emergent pathogens.

A resume of all the methods described in this review for drug repurposing is given in Table 3.

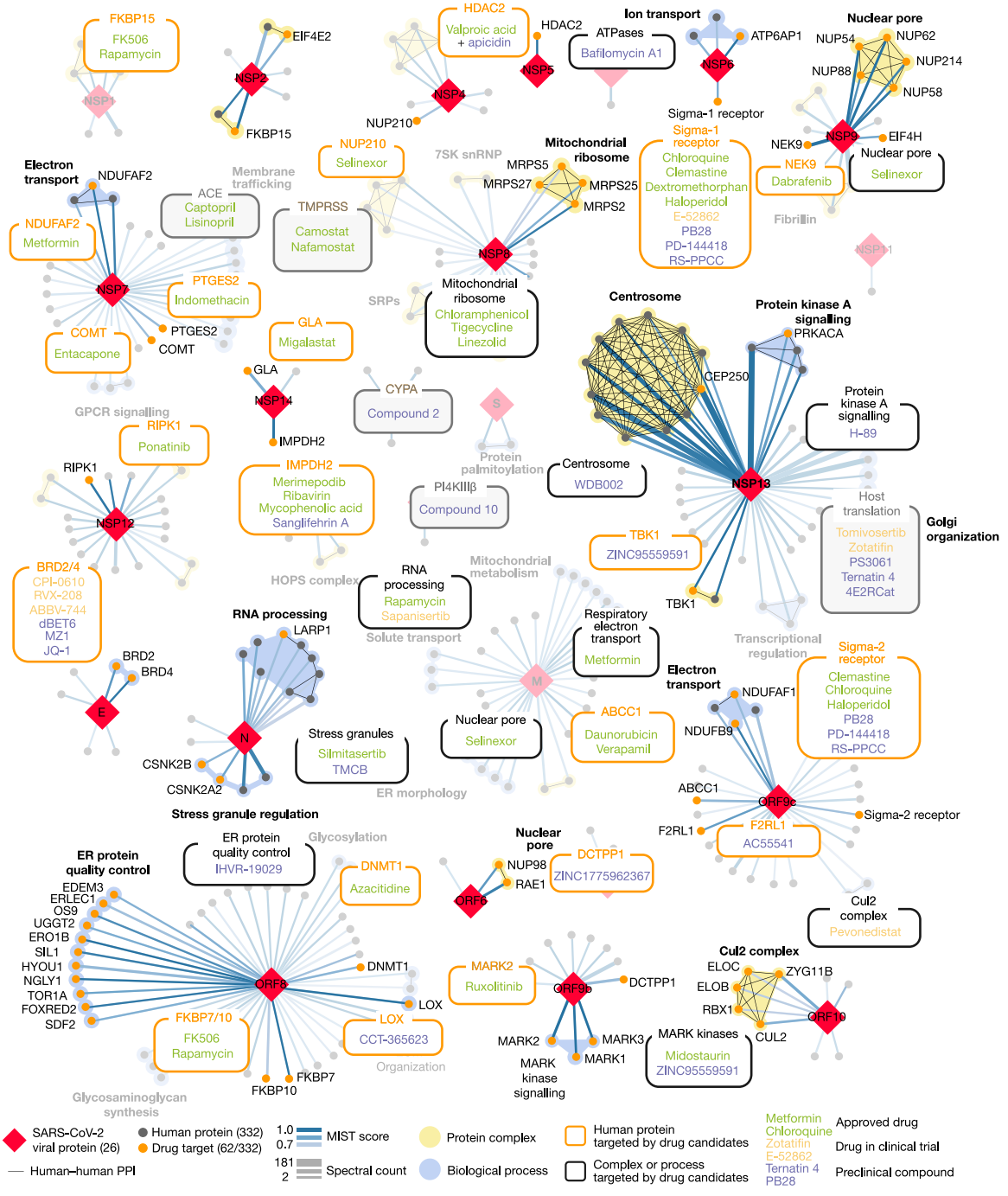


Fig. 3.8. Illustration of the PPIs of SARS-CoV-2 baits with approved drugs (green), clinical candidates (yellow), and preclinical candidates (purple) with experimental activities against the host proteins (white background) or previously known host factors (gray background). Figure provided by the authors.
Source: Reproduced from Gordon et al. [84] with permission.

4. Vaccines

A biological preparation which provides active acquired immunity to a particular infectious disease is known as a vaccine. The goal of a vaccine is to stimulate the immune system of the host to recognize the infectious agent and destroy it. It also should recognize and destroy any of the microorganisms associated with that infectious agent that it may

Table 3
Resume of the different methods described here for drug repurposing.

Authors	Target	Method	Examples of drugs identified	Ref.
Jin et al.	Mpro	virtual screening	cinanserin	[93]
Khan et al.	Mpro	molecular docking + molecular dynamics	remdesivir, saquinavir, darunavir	[97]
Tsuji	Mpro	virtual screening	sulfamethizole, sulfathiazole, kanamycin, droperidol	[98]
Hall & Ji	Mpro	in silico docking	zanamivir, indinavir, saquinavir, remdesivir	[99]
Joshi et al.	Mpro	virtual screening	d-viniferin, myricitrin, afzelin, hesperidin	[100]
Jiang et al.	nsp16	molecular docking	MK3207, rimegepant, entrectinib, bolazine, hesperidin	[115]
Sharma et al.	nsp16	homology modeling + virtual screening	saquinavir, indinavir	[116]
Khan et al.	nsp16	homology modeling + virtual screening	dolutegravir, bictegravir	[117]
de Oliveira et al.	S	molecular docking	ivermectine, quinupristin, acetyldigitoxin	[120]
Wei et al.	S	virtual screening	raltegravir, digotoxin	[122]
Villoutreix et al.	S	molecular docking	sulconazole	[123]
Borgio et al.	nsp13	homology modeling + virtual screening	vapreotide, atazanavir, hydroxychloroquine, lopinavir, saquinavir, indinavir, atazanavir	[125]
Hillen et al.	nsp12	molecular docking	saquinavir, tipranavir, lonafarnib, tegobuvir	[126]
Arya et al.	nsp3	homology modeling + virtual screening	biltricide, cinacalcet, procainamide, terbinafine	[129]
Wu et al.	several	homology modeling + virtual screening	ribavirin, valganciclovir, lymecycline, chlorhexidine, ceftibuten	[130]
Beck et al.	several	homology modeling + virtual screening	atazanavir, remdesivir, efavirenz, ritonavir	[131]
Gordon et al.	several	PPI network-based protein–drug association	zotatifin, ternatin-4, haloperidol, PB28, hydroxychloroquine	[84]
Gysi et al.	several	PPI network-based protein-drug association + AI	isoniazid, troleandomicyn, cilostazol, chloroquine	[135]
Zhou et al.	several	predicted PPI network-based protein-drug association	mesalazine, toremifene, eplerenone, paroxetine	[137]

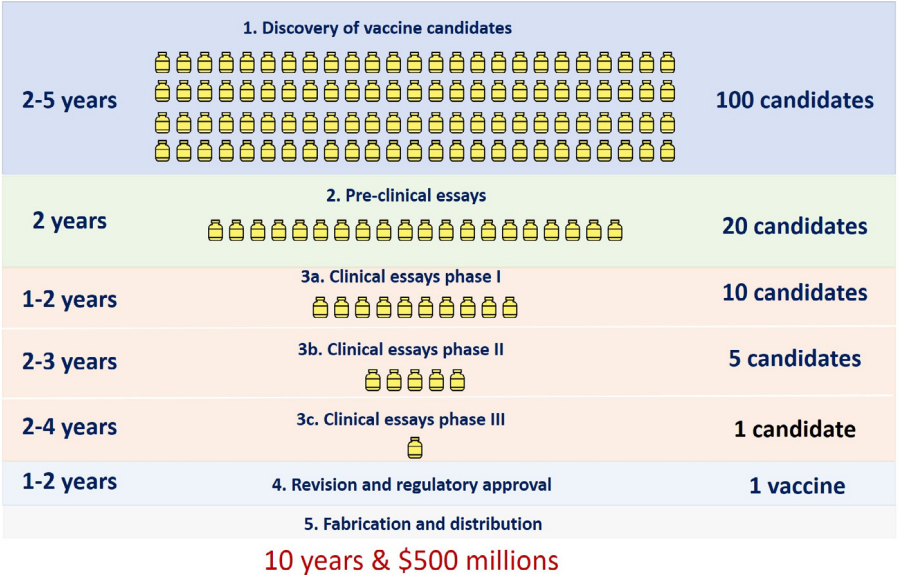


Fig. 4.1. Global scheme of vaccine development pipeline.

encounter in the future. A vaccine that prevents or ameliorates the effects of a future infection is known as prophylactic, while one that fights against a disease already in course is known as therapeutic [138]. To start we should mention here how the process for developing a new vaccine takes place. In Fig. 4.1 we illustrate the general process of discovering a vaccine showing the narrow funnel conducting from about 100 candidates to one single vaccine in a period of 10–12 years and a cost of $500 million.[14] Although other authors [138] give estimations of up to 16 years and up to one billion USD.

[14] The Figure is adapted from https://www.weforum.org/agenda/2020/06/vaccine-development-barriers-coronavirus/.

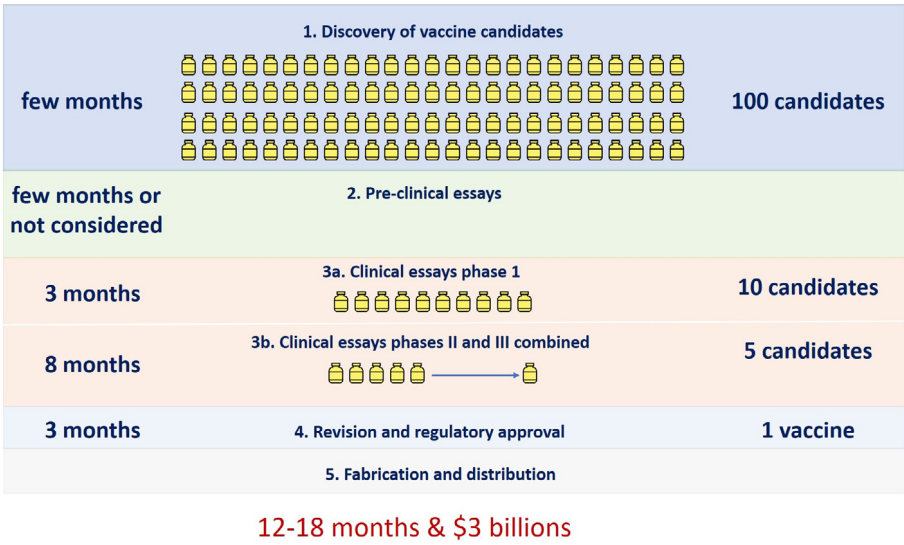


Fig. 4.2. Proposed scheme for anti-SARS CoV-2 vaccine development.

The specific timing and cost of a vaccine depend on the kind of vaccine we are talking about, because in general they are divided into 3 generations based on the process involved. Here, we will briefly describe the existing generations of vaccines currently in use [138].

(1) First-generation vaccines. Consist of the classic strategy of using either the live attenuated virus or the complete inactive virus. This strategy has the advantage of generating inherent immunogenicity and stimulating Toll-like receptors. However, they require extensive testing to confirm their safety in humans. In the case of coronavirus an additional problem is that an increase in infectivity has been observed after immunization.
(2) Second-generation vaccines. Consist of the use of subunits of the pathogen, particularly a complete protein on the surface of the pathogen that triggers the immune response against the pathogen. They include (i) subunit vaccines, (ii) conjugated vaccines, and (iii) recombinant vaccines. In this case, the spike proteins are excellent candidates in the case of SARS CoV. For example, the receptor binding domain of the SARS CoV-1 and SARS CoV-2 spike proteins share 80% similarities in their amino acids and bind to the same receptor, which are potentially useful characteristics for developing a vaccine based on that protein.
(3) Third generation of vaccines. Consist of nucleic acids from the pathogen. These vaccines build on the concept of DNA immunization that has been used successfully for the development of influenza vaccines. They exploit the immunogenic potential of plasmid containing a gene which encodes the antigen. Therefore they are also known as genetic vaccines.

In the emergency situation produced by the outbreak of COVID-19 a drastic modification of the global pipeline for vaccine development has been proposed [139–143]. It is resumed in Fig. 4.2 where the most relevant features are the reduction of the total timing to 12–18 months mainly by reducing the time used for the discovery of candidates, mainly due to the previous investigations done for SARS CoV-1 and MERS [144], the reduction of the time or even elimination of the pre-clinical essays, due to the same previous reason, and the merging of phases II and III of clinical essays. However, the cost will significantly rise to more than $3 billion. It should be remarked that for the case of COVID-19 we play with some advantages due to the fact that previous researches performed for SARS CoV-1 can be beneficial for the current development [144]. Researchers began the development of several SARS CoV-1 vaccines which were tested in animal models. These included recombinant vaccines based on the CoV spike proteins. Other vaccines were based on complete or attenuated viruses, and vector vaccines were also developed. Although most of these vaccines protected animals from the challenge of SARS-CoV-1, most of them did not induce sterilizing immunity. Only a small number of these vaccines reached Phase I of the clinical trials and the main reason was that research and development funding was cut due to the eradication of the virus in the human population. What experiences can be drawn from this episode? (1) that some of the vaccines in development against SARS-CoV-1 appeared safe and induced neutralizing antibodies; (2) that some monoclonal antibodies isolated from SARS-CoV-1 may also react against SARS-CoV-2, although since vaccines have not been developed they are not currently available; (3) that in some cases total virus-based vaccination resulted in complications, including lung damage, eosinophil infiltration, and liver damage from model animals; (4) that the time for the development of a vaccine should be dramatically shortened.

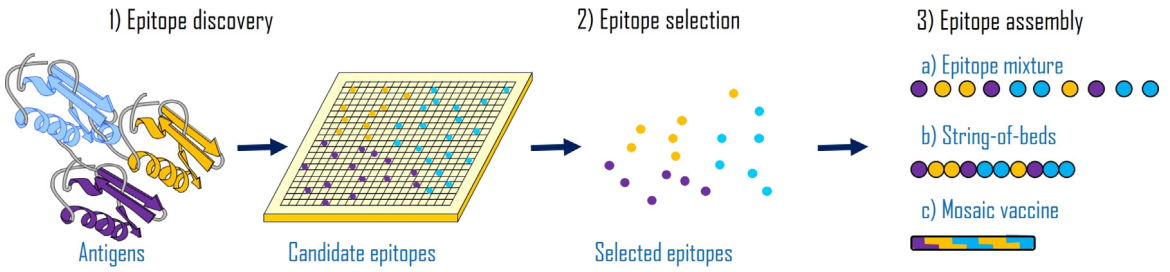


Fig. 4.3. The main steps of the rational vaccine design pipeline adapted from a diagram in [146] and made using Motifolio.

The use of this strategy has yielded several vaccines against SARS CoV-2 which are now in clinical trials (on June 29th 2020) and several others in preclinical testing.[15] In the case of SARS-CoV-2 vaccines, it must be taken into account that a highly vulnerable segment of the population is over 60 years of age. Unfortunately, this segment of the population typically responds somewhat worse to vaccination due to immunosenescence-aging of the immune system. Therefore, a strategy in the case of SARS CoV-2 would be the use of a fourth generation of vaccines: epitope vaccines.

4.1. Epitope vaccines

Epitopes are short sequences of amino acids belonging to a protein that can induce a more direct and powerful immune response than the one induced by the complete protein. Put simply, instead of using the whole virus (first generation of vaccines) or a complete virus protein (second generation of vaccines), let us use the part of this protein that binds directly to the antibody [145]. The advantages of epitope vaccines would be that: (i) they can be produced quickly and stored safely using available technology, (ii) they do not pose a risk of reversion to virulence since they do not contain infective material, and (iii) they can choose epitopes that take into account both the pathogenic and genetic variability of the population.

For the development of an epitope vaccine it is first necessary to discover the candidate epitopes, then select the necessary ones for vaccination and finally assemble them in the vaccine (see Fig. 4.3). The assembly stage is very important since the vaccines can consist of an epitope soup, called cocktail epitope vaccines, or they can be assembled in a linear fashion, called string-of-beds vaccines, or overlapping the epitopes with each other to give place to a mosaic vaccines (see Fig. 4.3). The problem of designing epitope vaccines can then be formulated mathematically as a combinatorial optimization problem (see further). An amino acid sequence is selected that must contain several epitopes and that must maximize the efficiency of antibodies against the virus. Since some epitopes appear more frequently than others in natural viral populations, there is a tendency to maximize the coverage of epitopes that appear in the vaccine by limiting their length. This should ensure that the most common epitopes are most likely to be included in the vaccine. To this end, various computational, statistical, and mathematical techniques have been used, such as genetic algorithms, linear programming, probabilistic algorithms, consensus methods, and, more recently, a unified epitope vaccine design formalism that simultaneously attacks all three stages has been developed. This method combines the use of graphs or weighted directed networks to represent the epitopes and their different forms of assembly with combinatorial optimization techniques that seek to maximize the immune response of the vaccine, imposing limits on the length of the artificial antigen built for the vaccine by the overlap of epitopes found in the virus (see further for details and references).

In 2014 a group of scientists from different institutions designed a vaccine using respiratory syncytial virus epitopes via computational protein design techniques [147]. They demonstrated that small conformationally and thermally stable scaffold proteins can be generated to emulate the structure of viral epitopes very well and induce potent virus neutralizing antibodies. The work by Correia et al. [147] is considered as a proof of principles for the design of vaccines based on epitopes. In general, the computational design of vaccines is framed within "immunoinformatics" as a branch of bioinformatics [148]. It includes the development of mathematical and computer techniques for the design of structure-based immunogens, the analysis of antibodies, the creation and management of databases and tools on cellular epitopes, prediction of T-cell epitopes, allergy, immunological genes, vaccinology in silico, among others.

4.2. Immunoinformatics and vaccinology for SARS CoV-2

4.2.1. Cell entry mechanism of SARS CoV-2

We have previously mentioned that the entry of SARS CoV-2 into human cells is facilitated by the spike (S) protein [149–151]. It is a glycoprotein consisting of 1282 residues, and divided into two regions known as S1 or residue binding domain (RBD) and region S2 [149], [152]. The RBD is formed by amino acids 1–685 (see Fig. 4.4 zoomed right

[15] The vaccines in clinical trial can be searched at the webpage of ClinicalTrials.org.

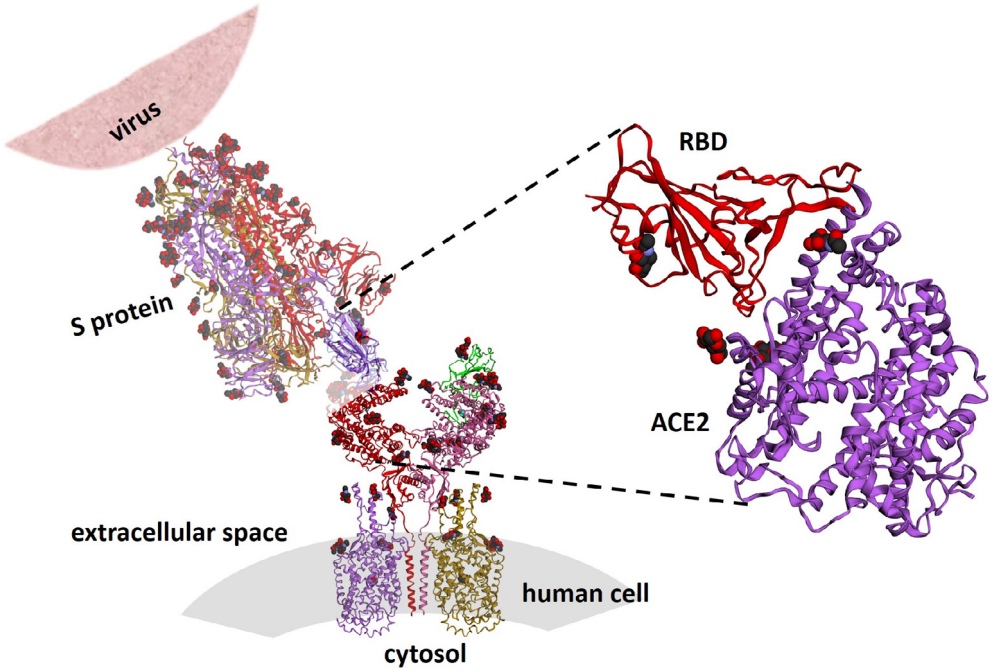


Fig. 4.4. Composition of the structures of SARS CoV-2 S protein, ACE2 receptor and neutral amino acid transporter B0AT1 forming the entry complex of the virus into the human cell.

part) and it is responsible of the binding to the cellular receptor, Angiotensin-converting enzyme 2 (ACE2). In addition, entry requires S protein priming by cellular proteases, which entails S protein cleavage at the S1/S2 site. The S2 region is formed by residues 686–1273, and mediates membrane fusion, a process previously explained in this review. In the process of protein priming the S protein requires TMPRSS2 (transmembrane serine protease 2) [153]. The 3D structure of SARS CoV-2 S trimer in the prefusion conformation was determined at 3.5-angstrom-resolution by using cryo-electron microscopy by Wrapp et al. [118]. The authors found that the RBD domain can be in two states: "up" or "down", but it predominantly is in the "up" conformation. In Fig. 4.4 (left part) we illustrate the structure of the SARS CoV-2 S trimer with the RBD in up conformation (PDB ID: 6VSB). They also provided an important differentiation of the S protein of SARS CoV-2 with that of SARS CoV-1. Namely, that (i) the new spike protein binds ACE2 with higher affinity than that of SARS CoV-1, and (ii) SARS CoV-1 RBD-specific monoclonal antibodies do not have appreciable binding to SARS CoV-2 S protein, which means that antibody cross-reactivity could be limited between the two RBDs. On the other hand, Yan et al. [154] determined the 3D structure of the human ACE2 receptor in the presence of the neutral amino acid transporter B0AT1. They considered the complex with or without the RBD of SARS-CoV-2 S protein. In Fig. 4.4 (left part) we show this complex with the RBD region (PDB ID: 6M17), which we have manually assembled with the S protein structure of SARS CoV-2 to present the whole complex of the interaction between the virus S protein and the human receptor at the cell surface. The zoomed interaction (Fig. 4.4, right part) corresponds to the crystal structure determined by Wang et al. [152] (PDB ID: 6LZG).

The previously described structural results are the basis for the search of epitopes in the S protein. This was demonstrated experimentally by Poh et al. [155] who used pools of overlapping linear B-cell peptides, and reported two IgG immunodominant regions on SARS CoV-2 S protein. These epitopes were recognized by sera from COVID-19 convalescent patients. Both epitopes are located at the surface of the S protein, one of them is located near the RBD and is specific to SARS CoV-2, and the other, which is at the fusion peptide, is potentially useful as a pan-SARS target. In another study, Barnes et al. [156] reported an epitope that blocks ACE2 receptor binding using a neutralizing monoclonal Fab-spike complex. Yuan et al. [157] determined the crystal structure of a neutralizing antibody previously isolated from a convalescent SARS patient. The antibody named CR3022 is in complex with the RBD of the S protein of SARS-CoV-2. They found that CR3022 targets a highly conserved epitope that enables cross-reactive binding between SARS-CoV-2 and SARS-CoV. In the next subsections we will describe some works reporting computational efforts to detect epitopes that may help in the design of vaccines against SARS CoV-2.

4.2.2. Immune response against SARS CoV-2

Here we briefly explain the mechanisms of immune response at molecular and cellular levels against SARS CoV-2. As we will see these are similar mechanisms to the ones generated by epitope vaccines. Therefore we use this subsection to

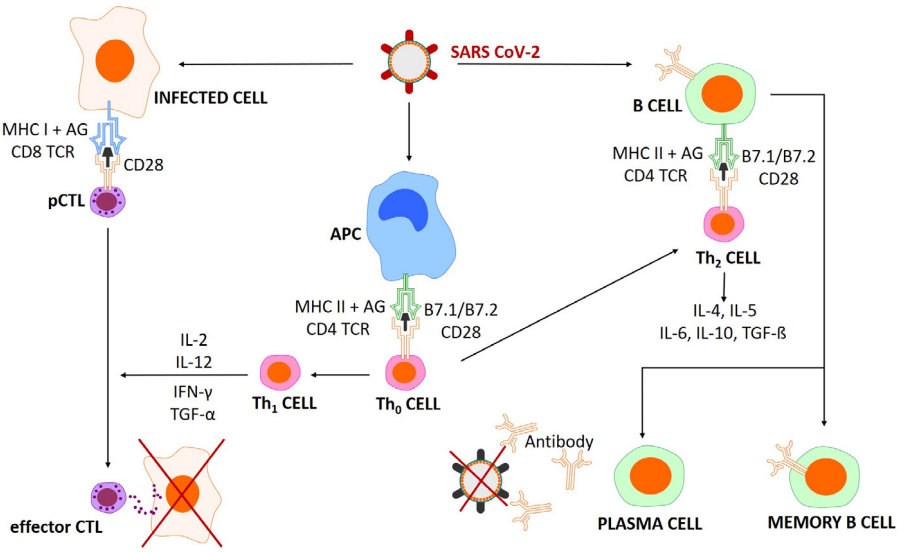


Fig. 4.5. Molecular and cellular mechanism of immune responses induced against SARS CoV-2. Graphics prepared with Motifolio.

introduce a few concepts that will be used in the next subsections of this review. When SARS CoV-2 enters the human cell, cytotoxic T lymphocytes (CTL), helper T-cell (Th) and B cells epitopes from some of its proteins can trigger an immune response from the infected cell. First, CD8+ precursor CTL (CD8+ pCTL) recognizes the complex of CTL antigen peptides bound to the major histocompatibility complex (MHC) class I molecules that are displayed by target cells through TCR. CD8+ refers to cytotoxic T cells accompanied by a glycoprotein called CD8. The T-cell receptor (TCR) is a protein complex found on the surface of T cells, or T lymphocytes, which is responsible for recognizing fragments of antigen as peptides bounded to MHC molecules. This first step is illustrated in Fig. 4.5 (left part).

In the central part of Fig. 4.5 we illustrate how antigen presenting cells (APC) tackle the epitopes generated by SARS CoV-2 and present the Th antigen peptides bound to MHC class II molecules to Th_0 cells. The cells Th_0 are differentiated into Th_1 and Th_2, where the first secrete interleukines (IL-2, IL-12) an interferon (IFN- γ) and transforming growth factor alpha (TGF- α) that stimulate CD8+ pCTL to generate effector CTL cells. This will kill the target cells as illustrated at the bottom of the left part of Fig. 4.5. The cells Th_2 (see right part of Fig. 4.5) recognize the Th epitope bound to MHC class II molecules that are presented by B cells. After activation these cells secrete IL-4, IL-5, IL-6, IL-10 and TFG- β to stimulate B-cell activation. B cells then proliferate and differentiate into plasma cells. The plasma cells secrete anti-SARS CoV-2 antibodies to kill the virus. Then, if a vaccine is created by containing CTL, Th and B cells epitopes of SARS CoV-2, the effective response against the virus created by this multi-epitope vaccine is similar to the one described in Fig. 4.5 [158].

4.2.3. Computational search of epitopes for SARS CoV-2 vaccines

Although in the previous subsection we have reported cases in which one or two epitopes have been found at the surface of S protein we should remember that a significant immune response in the body is rarely obtained by a single epitope [159]. In most cases this immune response is obtained by induction of CD4+ T cells to assist B cell expansion and differentiation, class switching, and maturation [160]. The use of computational techniques for helping in the design of vaccines has been reviewed by Poland et al. [161] where the terms "reverse vaccionology", "vaccionomics", "systems vaccionology", "structural vaccinology" and "vaccine informatics", together with "adversomics" are surveyed. More specific reviews for the use of computational tools for epitope vaccine design and evaluation have been published by He and Zhu [162] and by Khalili et al. [163].

As a first example we describe here the work of Kalita et al. [164] who designed a multi-peptide subunit-based epitope vaccine against COVID-19. Their recombinant vaccine consists of an adjuvant, cytotoxic T-lymphocyte (CTL), helper T-lymphocyte (HTL), and B-cell epitopes joined by linkers, which, according to the predictions, must be (i) non-toxic, (ii) non-allergenic, (iii) thermostable, and (iv) capable of eliciting a humoral and cell-mediated immune response. The first modeling tool used by these authors, after having the sequences of all proteins of SARS CoV-2, was for predicting the average antigenic propensity of each of these proteins. They used the antigenic peptides prediction tool,[16] which predicts those segments from within a protein sequence that are likely to be antigenic by eliciting an antibody response. It uses a method based on a table that reflects the occurrence of amino acid residues in experimentally known segmental epitopes. According to the search of Kalita et al. [164] the membrane glycoprotein, the S protein, and the nucleocapside protein

16 http://imed.med.ucm.es/Tools/antigenic.pl.

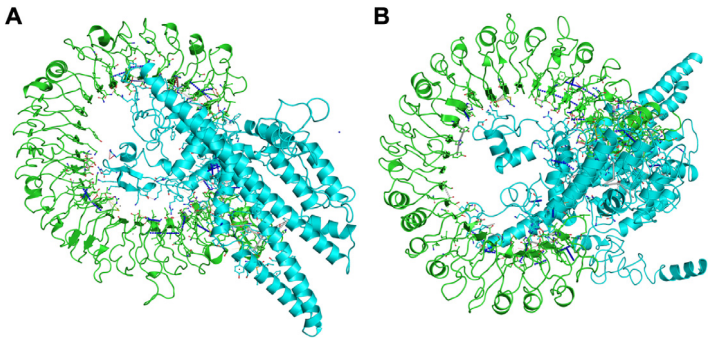


Fig. 4.6. Illustration of the key interactions obtained from the structure of TLR3 and vaccine complex, before (A) and after (B) molecular dynamics simulation. TLR3 receptor is shown in green color, and the vaccine is shown in cyan color in both panels.
Source: The figure was provided by Kalita et al. [164] and reproduced with permission from Elsevier.

display the highest order of antigenicity among all proteins in SARS CoV-2. We mention in passing that the structure of the nucleocapside protein has been recently determined by X-rays crystallography (PDB ID: 6M3M) [165]. Then, the methods and computational tools given below were used for the three proteins selected for the study:

- Immune Epitope Database[17]: For predicting the helper T-lymphocyte (HTL) epitopes;
- IFN epitope server[18]: For detecting capability to induce Th1 type immune response accompanied by IFN-Υ production;
- NetCTL1.2 server[19]: For predicting cytotoxic T-lymphocyte (CTL) epitopes;
- ABCPred server[20]: For identifying B-cell epitopes;
- ToxinPred module[21]: For predicting the toxic/non-toxic nature of all the epitopes.

Other computational tools can be found in Khalili et al. [163]. At this point the authors disposed of a series of (HTL specific, B-cell binding and CTL) epitopes from the three proteins considered. Therefore, they proceeded to construct the multi-epitope vaccine by adding adjuvant (human β-defensin 1) and linkers to separate the epitopes in vivo. The properties of the designed vaccine were then predicted by using the tools resumed below:

VaxiJen[22]: For predicting immunogenicity;
ANTIGENpro module of SCRATCH[23]; AllerTOP[24]; AlgPred Server[25]: For predicting allergenicity;
ProtParam[26]: For predicting physicochemical properties;
PSIPred 4.0 Protein Sequence Analysis Workbench[27]: For predicting secondary structure;
trRossetta[28]: For predicting tertiary structure;
PROCHEK v.3.5[29] and ProSA[30]: For model validation.

The modeling is finished by using molecular simulations of the 3D structure of the vaccine and of its receptor (TLR3) giving rise to the complex illustrated in Fig. 4.6.

Similar approaches were followed by ul Qamar et al. [166] and Bhattacharya et al. [167] who designed multi-epitopes vaccines against SARS CoV-2. The first considered seven antigenic proteins taken as targets, from which curiously the S protein was excluded due to low antigenicity, and selected several epitopes for building the vaccine. The second study, however, focuses only on the S protein from which the authors identified several B- and T-cell epitopes which then are assembled in a vaccine further optimized using molecular docking.

17 http://too73ls.iedb.org/mhcii/.
18 http://crdd.osdd.net/raghava/ifnepitope/.
19 http://www.cbs.dtu.dk/services/NetCTL/.
20 http://crdd.osdd.net/raghava/abcpred/.
21 http://crdd.osdd.net/raghava/toxinpred/multi_submit.php.
22 http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html.
23 http://scratch.proteomics.ics.uci.edu/.
24 http://www.ddg-pharmfac.net/AllerTOP/.
25 http://crdd.osdd.net/raghava/algpred/.
26 http://web.expasy.org/protparam/.
27 http://bioinf.cs.ucl.ac.uk/psipred/.
28 https://yanglab.nankai.edu.cn/trRosetta/.
29 https://servicesn.mbi.ucla.edu/PROCHECK.
30 https://prosa.services.came.sbg.ac.at/prosa.php.

The identification of specific epitopes based on the computational analysis of SARS CoV-2 proteins has received more attention. In one of these studies Grifoni et al. [168] used the Immune Epitope Database and Analysis Resource (IEDB) mentioned before to identify epitopes in coronaviruses. Due to the high homology of SARS CoV-2 to the SARS-CoV-1 virus and the availability of epitopes for SARS CoV-1 they used information about the last for identifying potential B and T cell epitopes for SARS-CoV-2. Interestingly, they used independent identification of the same regions by two different approaches and found that a few of the epitopes identified coincide, reflecting a high probability that these regions are potential targets for immune recognition of SARS-CoV-2. It is important to mention that among the IEDB inventory of epitopes for SARS CoV-1 proteins about 65% of B cell ones and 47% of T cell ones comes from the Spike protein. Then, from the 10 dominant SARS CoV-2 B cell epitopes identified, 5 are from the S protein, 2 from the membrane protein and 3 from the nucleocapside one. In the case of SARS CoV-2 T cells they identified 45 epitopes, 23 of which are from S protein, 10 from the nucleocapside, 7 from Orf1ab, 3 from the membrane protein and 2 from Orf3a.

Zheng and Song [169] computed sequence-based antibody epitope scores in spike proteins of MERS-CoV, SARS-CoV, and SARS-CoV-2. They found that the SARS-CoV-2 had significantly lower antibody epitope score compared with MERS-CoV and significantly higher antibody epitope score compared with SARS-CoV-1 which indicates that the spike proteins have significantly variable antigenicity. Their results show that although SARS CoV-1 and SARS CoV-2 have high homology it is necessary to develop new antibodies and vaccines specific for SARS-CoV-2. They discovered novel and high-score antibody epitopes for SARSCoV-2 S protein with enough potency and specificity as for developing anti-SARS CoV-2 vaccines. Campbell et al. [170] computationally searched epitopes predicted to bind any class I HLA protein across the entire SARS-CoV-2 proteome. They computed the predicted binding affinities between 9-mer peptides derived from the annotated SARS-CoV-2 peptidome across 9,360 MHC class I HLA-A, -B, and -C alleles. This resulted in 6748 unique combinations of peptides and HLA alleles with a predicted binding affinity of less than 500 nM. From this list 1103 were unique peptides and 1022 were HLA alleles, spanning 11 annotated superfamilies. The complete dataset with their results is publicly available.[31] Finally, Kiyotani et al. [171] performed a comprehensive bioinformatic screening of potential epitopes from the SARS-CoV-2 sequences for HLAs commonly present in the Japanese population. They first found 2013 and 1399 potential epitopes with predicted high affinity to HLA classes I and II, respectively. These epitopes are located across the spike, envelope, membrane, nucleocapsid and in six open reading frames proteins. From these epitopes the authors found that 781 HLA-class I and 418 HLA-class II have high homologies with SARS CoV-1. Their study was completed by identifying 2 HLA-class I epitopes covering 83.8% of the Japanese population which may be useful for designing widely-available vaccine against SARS-CoV-2. In another study Prachar et al. [172] identified 174 SARS-CoV-2 epitopes with high prediction binding scores, validated to bind stably to 11 HLA allotypes. Also, Ahmed et al. [173] reported a set of B cell and T cell epitopes derived from the spike and nucleocapsid proteins. For the T cell epitopes, the authors performed a population coverage analysis of the associated MHC alleles and proposed a set of epitopes that is estimated to provide broad coverage globally.

4.3. Mathematical modeling of epitope vaccines

The design of epitope vaccines consists of three main stages: (i) the discovery of potential epitopes; (ii) the selection of a subset of epitopes to be included in the vaccine; and (iii) the assembly of epitopes into the vaccine. We have seen in the previous subsections that there are several bioinformatic [148] tools to discover potential epitopes and to select a reduced set of them meeting several criteria. In order to remark the importance of the last stage we should recall that delivering all a mixture of separate epitopes does not induce immune response and it is not to be considered as a vaccine. Therefore, the way in which the selected epitopes are assembled is vital for the efficiency of an epitope vaccine.

An apparently simple way of assembling the selected epitopes is by concatenating them in the form of a sequence in the form of string-of-beds (see Fig. 4.3). A string-of-beds vaccine is then a long polypeptide whose efficacy depends on its capacity of recovering the majority of epitopes to be recognized by human leukocyte antigen (HLA) molecules. Consequently, the order in which these epitopes are arranged in the string matters as it has been shown by Cornet et al. [174]. The recovery of epitopes can be improved by including spacers sequences between epitopes as it has been shown in the previous subsection. However, as recognized by Schubert and Kohlbacher [175] there are a few problems that can emerge, such as:

- unfavorable ordering of epitopes which may result in miscleaved peptides and ineffective vaccines;
- formation of neoepitopes, which can have detrimental effects. This situation can be aggravated by using longer spacers;
- combinatorial explosion of the number of possible arrangements, which makes experimentally untreatable the problem.

We should remember that the possible number of arrangements for N epitopes is $N!$. Thus, with only 10 epitopes we already have 3,628,800 of combinations to try in the wet lab. The mathematical problem of optimizing the three kinds of epitopes vaccines can be formulated as follows [146]:

- ***Epitope mixture vaccines***: seeks to find a subset P of k epitopes that together have the highest chance of invoking an effective immune response $I(P)$;

[31] gs://picicovid19- data-resources/mhci/peptide_predictions.

- ***String-of-beads vaccines***: seeks to find a polypeptide comprised of k concatenated epitopes that simultaneously maximize the vaccine efficacy $I(P)$ and the recovery likelihood of each epitope by the proteasome;
- ***Mosaic vaccines***: seeks to constructs an artificial antigen P of fixed length h comprised of potentially overlapping epitopes with maximal efficacy $I(P)$.

The assembly as string-of-beds is possibly the most popular one and there are several approaches to optimize the corresponding arrangement of strings. A well-known mathematical formalism for epitope selection and ordering was proposed by Toussaint et al. [176]. They consider a process that balances the constraints on: (a) optimal choice of target antigens, (b) optimal choice of highly conserved epitopes, (c) maximum coverage of the target population, and (d) the proper ordering of the epitopes in the final vaccine. Then, they defined the effectivity of an epitope with respect to a set of MHC alleles as a weighted sum of the epitopes' presentation probability scores, which are composed of cleavage probability and MHC binding affinity. That is,

$$\text{Effectivity}\,(e, A) = \sum P_{mhc}\,(a)\,p_{cl}\,(e)\,b\,(e, a)\,, \tag{4.1}$$

where e is the epitope of interest, A is the set of MHC alleles, $P_{mhc}(a)$ is the probability of allele a in the target population, $p_{cl}\,(e)$ is the cleavage probability of epitope e, and $b\,(e, a)$ is the binding affinity of epitope e to the gene product of MHC allele a. Then, the selection of the epitopes to form the vaccine is made by using an integer linear program (ILP) to determine the optimal set of epitopes, based on three parameters: (i) the protein size of immune repertoire scores, (ii) epitope effectivities, and (iii) epitope conservation. Finally, the assembly is performed as follows. Let us consider the epitopes as the nodes of a directed weighted graph where the edge (i, j) represents the number of unwanted junctional peptides likely to be cleaved when i is the N-terminal neighbor of j or vice versa such that the more unwanted peptides, the higher the weight. Let us introduce a dummy epitope corresponding to the N- and the C-terminus of the polypeptide. Then an ILP was implemented to solve the traveling salesman problem[32] on this graph.

The strategy followed by Schubert and Kohlbacher [175] is to design a string-of-beds such that it: (a) maximizes the recovery of the epitopes, while (b) minimizing the production of undesired neo-epitopes. They also created weighted directed fully connected graphs where the negative cleavage scores represent the weights of the edges between epitopes pairs. Then, they used Toussaint et al. [176] approach based on traveling salesman problem on this graph also by adding a node that represents the N- and C-termini of the vaccine and connecting it with all other nodes with zero edge weights. One of the main additions of this work is to consider optimal spacers of length k connecting two epitopes. Therefore, their method yields string-of-bead vaccines with flexible spacer lengths. This was shown to increase the predicted epitope recovery rate 5 times, while reducing the immunogenicity from neo-epitopes by 44% compared to designs without spacers.

Another variation is presented by Epigraph,[33] which is a developed algorithm enabling to maximize the potential epitope coverage for a diverse pathogen population and so the design of single or multiantigen vaccines. The problem is formulated again on graphs [177], but this time the edges of the graphs represent the overlapping between two epitopes. The candidate antigens are represented as walks[34] that traverse this graph (notice that although they call it paths, they allow for repetition of nodes, so it is more correct to call them walks). In a more recent variation of the theme, Dorigatti and Schubert [178] proposed to solve the problem of (1) selecting the best epitopes eliciting the strongest possible immune response, and (2) arranging and linking the selected epitopes through short spacer sequences to string-of-beads vaccines so as to increase the recovery likelihood of each epitope during antigen processing at the same time. Here again they based their approach on linear programming and solves both design steps simultaneously. They allowed to weigh the selection of a set of epitopes that have great immunogenic potential against their assembly into a string-of-beads construct that provides a high chance of recovery. The authors followed Toussaint et al. [176] and defined an overall contribution of an epitope to the vaccine immunogenicity as the weighted average of the log-transformed HLA binding strengths over a specified set of HLA alleles.

An important parameter to be taken into account during vaccine design is the vaccine length, which influences the production and transfusion of the vaccine. The problem was approached by Vider-Shalit et al. [179] from a computational perspective by calculating the peptide cleavage probability, transfer through TAP and MHC binding for a large number of HLA alleles. Then, the resulting peptide libraries were pruned for peptides that are not conserved or are too similar to self peptides. They used a genetic algorithm to produce an optimal protein composed of peptides from this list properly ordered for cleavage by minimizing the length of the vaccine.

The problem of epitope assembly has also been considered from a combinatorial optimization perspective, for instance by Martínez et al. [180,181]. Their approach can be resumed as follows. Given two sets of strings, a set of host strings, which models a set of instances of a protein, i.e., amino acid sequences of the protein for a given pathogen, and a set of target strings, which models a set of epitopes, a λ-superstring is defined as a string that models a candidate vaccine containing, as substrings, at least λ target strings from each host string. This means that the vaccine covers at least λ

[32] The traveling salesman problem is the archetypal problem in combinatorial optimization. ore information can be found in: E. Lawler, D. Shmoys, A. Kan, and J. Lenstra, The TravelingSalesman Problem (John Wiley & Sons, Incorporated, 1985).

[33] https://www.hiv.lanl.gov/content/sequence/EPIGRAPH/help.html.

[34] A walk in a graph is a sequence of (not necessarily different) consecutive vertices and edges. A walk is closed if the starting and ending vertices coincide. A path is a walk in which there is repetition of neither vertices nor edges.

epitopes in each patient. The associated optimization problem was to find a λ-superstring of minimum length, which means to find a candidate vaccine as short as possible.

A recent step forward in this combinatorial strategy was given by Dorigatti and Schubert [146] who presented a unifying formalism of the general epitope vaccine design problem. They considered all phases of the design process simultaneously and combined all prevalent design principles. The problem is finally formulated as an ILP which guarantees optimality of the designs. Their approach generalizes the optimal design of the three kinds of epitope vaccines illustrated in Fig. 4.3, whose design principles were stated before. The generalized scheme proposed by Dorigatti and Schubert [146] (DG) is then formulated as follow.

- ***DG generalized epitopes vaccine design***: Let $G(V, E, w)$ be a weighted, directed graph where the vertices V represent the epitopes and the weight $w(\cdot)$ of the edges E determine the design of the vaccine. An artificial node s representing the N- and C-terminus of the vaccine is added, such that it connects to every vertex $v \in V$ such that $w(e_{sv}) = a$ and $w(e_{vs}) = b$, with design-dependent weights $a, b \in \mathbb{R}$. The optimal vaccine in $G\left(\tilde{V}, \tilde{E}, w\right)$, with $\tilde{V} = V \cup \{s\}$ and $\tilde{E} = E \cup \{(s, v), (v, s) \,|\, v \in V\}$ it is needed to seek n disjoint subsets $P_1, \ldots, P_n \subseteq \tilde{V}$, each of size at most k, that together maximize the vaccine's immunogenicity $I : 2^V \to \mathbb{R}$, and whose simple tours $H(P_1), \ldots, H(P_n)$, start and end at $s \in \tilde{V}$ and weigh at most $h \in \mathbb{R}$, where the term simple tour refers to a closed walk with no repeated vertices except for s.

According to [146] this generalized approach makes it possible to explore new regions of the vaccine design space, analyze the trade-offs between the design phases, and balance the many requirements of vaccines.

5. Discussion

"All models are wrong, but some are useful" is a common saying in statistics, which has been heard many times during these days of the COVID-19 pandemic. This aphorism, generally attributed to the statistician George Box, expresses the fact that models necessarily have to make assumptions that simplify the reality. A cow is not spherical, although the volume of that spherical approximation can approach the reality under certain circumstances. The question is then to express clearly and explicitly which are the assumptions and limitations of the model, such that we can understand to which particular problems it can be applied. Under this paradigm, Box's saying can be turned around to say that "All models are right, but many are useless", namely because they are applied outside their proper scope or in a wrong way. Although this is true to many physical models we circumscribe ourselves here to those used to analyze and mitigate the impacts of a global pandemic.

In the situation of a global pandemic, in "war times" as it has been called by Vespignani [10], the impact of models goes beyond academia, and plays a fundamental role (or at least such a role is expected to be played) in political decisions, and in general public information. Therefore, the modeler should know the fact that her interlocutor is not necessarily aware of the scope and limitations of the model used. We have seen here that two global groups of models are in use for predicting epidemiological variables of COVID-19. One includes all compartmental models, from simple SIR to cumbersome models including many compartments, mobility, spatial, age, and other kind of data. The other group encloses mainly those statistical approaches based on data-driven methodologies as well as others with more limited scope. In general, most of the models presented here to discuss epidemiological variables do an excellent job in reproducing past data of the epidemic evolution. The problem arises when such models are used to predict future events. Several of the results discussed in this Review simply contain a curve indicating a unique trajectory of the epidemic in future times. This false impression of uniqueness given by deterministic epidemiological models has been criticized in the scientific literature (see [72]). The inference made by many authors that because their model reproduces well the past trajectory of the disease, it will reproduce well future outcomes is simply not correct. This is contrasting with the results illustrated in Fig. 2.19, where the confidence interval indicates that completely opposed outcomes can be reached if a prediction is made 40–60 days ahead from the last day of existing data. Opportunistic criticisms has then emerged against these (and other) models, and Box's saying has been heard truncated only at its first part, i.e., "all models are wrong". What is then recommended? Of course, we can and should continue using deterministic epidemiological models. But, if a serious job is going to be done with these models, then the modeler should:

(1) implement and report predictive posterior intervals for the model;
(2) make predictions only up to the point in which the lower and upper predictive intervals do not contradict each other, e.g., avoid the situations where one bound predicts exponential growth on the other a control of the epidemic;
(3) update systematically the model, such that the predictive horizon is increased and the capacity of predicting relatively short-time events increases the confidence in the model.

It should be reminded once again that the predictability of these models is not determined, although it is influenced by, the quality of the data and the number of compartments used in the model. It is an inherent property of exponential growing processes like the epidemic growth. Thus, the modeler should be aware that increasing the number of compartments and variables could make the model more susceptible to those initial conditions.

The other side of the coin can be illustrated by the statistical approaches yielding likelihoods of different scenarios. The confidence intervals given for the predictions of these models can be so wide that they range from assumable to catastrophic scenarios. These models are strongly dependent on the data existing to calibrate the epidemiological variables. In the current pandemic scenario some of these data, particularly when there has been lack of testing on population, have been scarce, and the modelers have to make reasonable assumptions (see [182]). It has also been remarked that such wide intervals projected, "limit predictability to the near future" [72]. Here again, the solution goes through the points (2) and (3) recommended before for the case of compartment models. In this case we can observe the dramatic improvement obtained by the application of this strategy (or something similar) for the case of the model used by Imperial College. In their Report 13, dated on March 30th, the prediction of the model for the percentage of total population infected (mean [95% credible interval]) for Spain was 15%, with a margin between 3.7% and 41% of the population [71]. This produced a great alarm in the country because it could represent, in the upper interval, more than 19 million infected individuals, while the lower bound represented less than 2 million. The same prediction, reported in a Nature paper accepted on May 22th, predicted for Spain (as of 4th May) 5.5% of the population infected with a range between 4.4% and 7.0% [183], which was a realistic scenario.

The second big area covered by this review is that of drug repurposing for targeting SARS CoV-2. This is a subfield of drug design, in which the researcher already dispose a list of candidate molecules, which are existing drugs, to be interrogated against one or several pharmacological targets in the virus. The great difference between modeling in this area and in that of epidemiology is that here the modeler can make controlled experiments and clinical trials to verify the results of her predictions. Evidently, we can differentiate between two different stages in the research for drug repurposing in the emergency of a pandemic like COVID-19. The first is in the early stage of the epidemic when the 3D structures of none of the proteins of the pathogen have been determined. In this case the use of many available bioinformatics tools for constructing protein 3D structure from homologies with existing ones is very recommendable and have proved useful in several previous instances. It should be remarked here that modelers in this area are trained to make such predictions which are then verified with the 3D structure of proteins. One example is the D3R Grand Challenge[35] competition, where the best existing models for drug design are identified. The competition is based on a few datasets, each comprising one protein and multiple ligands with measured affinity data, and with protein–ligand co-crystal structures for at least a subset of the ligands. The goal is to identify not only the best docking algorithms and scoring functions, but also to identify weak points in the existing ones. These methods and algorithms are also useful when the structure of the target is known, which evidently facilitates the work of the modeler. What is necessary in this area is a better collaboration between existing groups of modelers and those that can test the output of their virtual screenings. In other words, a link between the *in silico* and the *in vitro* worlds to approach more quickly to the *in vivo* universe.

Other areas of development of drug repurposing are younger and maybe least developed than the classical virtual screening by molecular docking. This includes many different applications of network theory to signature matching, genome-wide association, investigation of biological pathways, and others. More attention is needed for instance to the predictability of the existing methods for identification of protein targeted by a given pathogen. Attention to potential combinations between these new approaches and those existing for drug design and repurposing is also important. For instance, the investigation of ligand–protein interactions from a network perspective seems to reveal aspects not revealed by other approaches (see [108,112]).

The last area that we have reviewed here is that of vaccinomics and immunoinformatics, which is a promising, relatively new area of interdisciplinary research. This area is still young and many researchers in the physics and mathematics communities are not aware that their researches can have an important impact in accelerating the development of new vaccines in the future. Here we have make emphasis in epitope vaccines due to two main reasons. The first is that these are potentially the vaccines of the future, because by selecting an appropriate groups of epitopes we can address differences in the population due to ethnicity, age and sex. Also it is possible to identify epitopes that account for potential mutations of the virus that protect the population against future re-emergence of the same pathogen. The second is that the stage of epitope assembly into a vaccine represents a combinatorial challenge where different mathematical methods used in statistical physics can provide appropriate answers. We have seen here that the use of graphs/networks and combinatorial optimization are frequently used for that purpose. But we are optimistic that opening the scope of approaches to other physico-mathematical tools will be of benefit in vaccionomics and immunoinformatics.

All in all, we consider that modeling tools of a wide range is not only important but also somehow vital for our survival in front of this and future epidemics. We have walked a long way since the pioneering works that developed the first compartment models or implemented the first drug design approaches. We still need to improve our tools to tackle more complex challenges coming from emerging pathogens, and we hope that this review contributes to identify where we are, and what we still need to do to be well prepared the next time.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

[35] See the webpage of the competition t: https://drugdesigndata.org/about/grand-challenge.

Acknowledgments

The author is indebted to A. Aleta, A. Arenas, M. Bouhaddou, J. Cuesta, S. Eikenberry, J. Gómez-Gardeñes, E. Kuhl, C. Lock Yue, P. Van Mieghem, Y. Moreno, M. Perc, T. Tripathi, P. Wang, and G. Wei-Wei for sharing figures of their works to be included in the current review.

References

[1] V.C. Cheng, S.K. Lau, P.C. Woo, K.Y. Yuen, Severe acute respiratory syndrome coronavirus as an agent of emerging and reemerging infection, Clin. Microbiol. Rev. 20 (4) (2007) 660–694.
[2] M. Richard, A. Kok, D. de Meulder, T.M. Bestebroer, M.M. Lamers, N.M. Okba, M.F. van Vlissingen, B. Rockx, B.L. Haagmans, M.P. Koopmans, et al., SARS-CoV-2 is transmitted via contact and via the air between ferrets., Nat. Commun. 11 (2020) 3496.
[3] Y.M. Bar-On, A. Flamholz, R. Phillips, R. Milo, Science forum: SARS-CoV-2 (COVID-19) by the numbers, Elife 9 (2020) e57309.
[4] H.A. Rothan, S.N. Byrareddy, The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak, J. Autoimmun. 109 (2020) 102433.
[5] C. Vannabouathong, T. Devji, S. Ekhtiari, Y. Chang, S.A. Phillips, M. Zhu, Z. Chagla, C. Main, M. Bhandari, Novel coronavirus COVID-19: current evidence and evolving strategies, J. Bone Joint Surg. 102 (9) (2020) 734.
[6] S.P. Layne, J.M. Hyman, D.M. Morens, J.K. Taubenberger, New coronavirus outbreak: Framing questions for pandemic prevention, Sci. Transl. Med. 12 (534) (2020) eabb1469.
[7] A. Kucharski, Calculating virus spread, New Sci. 367 (6477) (2020) 23.
[8] J. Cohen, New coronavirus threat galvanizes scientists, Science 245 (3270) (2020) 492–493.
[9] Z. Cang, L. Mu, G.-W. Wei, Representability of algebraic topology for biomolecules in machine learning based scoring and virtual screening, PLoS Comput. Biol. 14 (1) (2018) e1005929.
[10] A. Vespignani, H. Tian, C. Dye, J.O. Lloyd-Smith, R.M. Eggo, M. Shrestha, S.V. Scarpino, B. Gutierrez, M.U. Kraemer, J. Wu, et al., Modelling COVID-19, Nat. Rev. Phys. 2 (2020) 279–281.
[11] W.O. Kermack, A.G. McKendrick, A contribution to the mathematical theory of epidemics, Proc. R. Soc. Lond. Ser. A 115 (772) (1927) 700–721.
[12] Y. Liu, A.A. Gayle, A. Wilder-Smith, J. Rocklöv, The reproductive number of COVID-19 is higher compared to SARS coronavirus, J. Travel Med. 27 (2) (2020) taaa021.
[13] A. Cori, N.M. Ferguson, C. Fraser, S. Cauchemez, A new framework and software to estimate time-varying reproduction numbers during epidemics, Am. J. Epidemiol. 178 (9) (2013) 1505–1512.
[14] M. Kivelä, A. Arenas, M. Barthelemy, J.P. Gleeson, Y. Moreno, M.A. Porter, Multilayer networks, J. Complex Netw. 2 (3) (2014) 203–271.
[15] S. Boccaletti, G. Bianconi, R. Criado, C.I. Del Genio, J. Gómez-Gardenes, M. Romance, I. Sendina-Nadal, Z. Wang, M. Zanin, The structure and dynamics of multilayer networks, Phys. Rep. 544 (1) (2014) 1–122.
[16] W. Mei, S. Mohagheghi, S. Zampieri, F. Bullo, On the dynamics of deterministic epidemic propagation over networks, Annu. Rev. Control 44 (2017) 116–128.
[17] M. D'Arienzo, A. Coniglio, Assessment of the SARS-CoV-2 basic reproduction number, R0, based on the early phase of COVID-19 outbreak in Italy, Biosaf. Health 2 (2) (2020) 57–59.
[18] J. Wangping, H. Ke, S. Yang, C. Wenzhe, W. Shengshu, Y. Shanshan, W. Jianwei, K. Fuyin, T. Penggang, L. Jing, et al., Extended SIR prediction of the epidemics trend of COVID-19 in Italy and compared with hunan, China, Front. Med. 7 (2020) 169.
[19] C. You, Y. Deng, W. Hu, J. Sun, Q. Lin, F. Zhou, C.H. Pang, Y. Zhang, Z. Chen, X.-H. Zhou, Estimation of the time-varying reproduction number of COVID-19 outbreak in China, Int. J. Hyg. Environ. Health 228 (2020) 113555.
[20] W.C. Roda, M.B. Varughese, D. Han, M.Y. Li, Why is it difficult to accurately predict the COVID-19 epidemic? Infect. Dis. Model. 5 (2020) 271–281.
[21] D. Fanelli, F. Piazza, Analysis and forecast of COVID-19 spreading in China, Italy and France, Chaos Solitons Fractals 134 (2020) 109761.
[22] C. Anastassopoulou, L. Russo, A. Tsakris, C. Siettos, Data-based analysis, modelling and forecasting of the COVID-19 outbreak, PLoS One 15 (3) (2020) e0230405.
[23] B. Prasse, M.A. Achterberg, L. Ma, P. Van Mieghem, Network-based prediction of the 2019-nCoV epidemic outbreak in the chinese province hubei, 2020, arXiv preprint arXiv:2002.04482.
[24] J. Wallinga, M. Lipsitch, How generation intervals shape the relationship between growth rates and reproductive numbers, Proc. R. Soc. B 274 (1609) (2007) 599–604.
[25] T. Zhou, Q. Liu, Z. Yang, J. Liao, K. Yang, W. Bai, X. Lu, W. Zhang, Preliminary prediction of the basic reproduction number of the Wuhan novel coronavirus 2019-nCoV, J. Evidence-Based Med. 13 (1) (2020) 3–7.
[26] Y. Fang, Y. Nie, M. Penny, Transmission dynamics of the COVID-19 outbreak and effectiveness of government interventions: A data-driven analysis, J. Med. Virol. 92 (6) (2020) 645–659.
[27] C. Hou, J. Chen, Y. Zhou, L. Hua, J. Yuan, S. He, Y. Guo, S. Zhang, Q. Jia, C. Zhao, et al., The effectiveness of quarantine of Wuhan city against the Corona Virus Disease 2019 (COVID-19): A well-mixed SEIR model analysis, J. Med. Virol. 92 (2020) 841–848.
[28] G.J. Fox, J.M. Trauer, E. McBryde, Modelling the impact of COVID-19 upon intensive care services in New South Wales, Med. J. Aust. (2020) Preprint, https://www.mja.com.au/system/files/2020-03/Preprint.
[29] N. Ferguson, D. Laydon, G. Nedjati Gilani, N. Imai, K. Ainslie, M. Baguelin, S. Bhatia, A. Boonyasiri, Z. Cucunuba Perez, G. Cuomo-Dannenburg, et al., Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand, 2020.
[30] A. Grant, Dynamics of COVID-19 epidemics: SEIR models underestimate peak infection rates and overestimate epidemic duration, 2020, medRxiv, https://www.medrxiv.org/content/medrxiv/early/2020/04/06/2020.04.02.20050674.full.pdf.
[31] M. Small, D. Cavanagh, Modelling strong control measures for epidemic propagation with networks–a COVID-19 case study, 2020, arXiv preprint arXiv:2004.10396.
[32] K. Linka, M. Peirlinck, F. Sahli Costabal, E. Kuhl, Outbreak dynamics of COVID-19 in europe and the effect of travel restrictions, Comput. Methods Biomech. Biomed. Eng. (2020) http://dx.doi.org/10.1080/10255842.2020.1759560, in press.
[33] N.N. Chung, L.Y. Chew, Modelling Singapore COVID-19 pandemic with a SEIR multiplex network model, 2020, medRxiv, http://dx.doi.org/10.1101/2020.05.31.20118372.
[34] H. Tian, Y. Liu, Y. Li, C.-H. Wu, B. Chen, M.U. Kraemer, B. Li, J. Cai, B. Xu, Q. Yang, et al., An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China, Science 368 (6491) (2020) 638–642.
[35] C. Manchein, E.L. Brugnago, R.M. da Silva, C.F. Mendes, M.W. Beims, Strong correlations between power-law growth of COVID-19 in four continents and the inefficiency of soft quarantine strategies, Chaos 30 (4) (2020) 041102.

[36] M. Peirlinck, K. Linka, F.S. Costabal, E. Kuhl, Outbreak dynamics of COVID-19 in China and the United States, Biomech. Model. Mechanobiol. (2020) http://dx.doi.org/10.1007/s10237-020-01332-5, in press.

[37] A.J. Kucharski, T.W. Russell, C. Diamond, Y. Liu, J. Edmunds, S. Funk, R.M. Eggo, F. Sun, M. Jit, J.D. Munday, et al., Early dynamics of transmission and control of COVID-19: a mathematical modelling study, Lancet Infect. Dis. 20 (5) (2020) 553–558.

[38] P. Block, M. Hoffman, I.J. Raabe, J.B. Dowd, C. Rahal, R. Kashyap, M.C. Mills, Social network-based distancing strategies to flatten the COVID-19 curve in a post-lockdown world, Nat. Hum. Behav. 4 (2020) 588–596.

[39] M.J. Keeling, T. Hollingsworth, J.M. Read, The efficacy of contact tracing for the containment of the 2019 novel coronavirus (COVID-19), 2020, medRxiv, http://dx.doi.org/10.1101/2020.02.14.20023036.

[40] A. Aleta, Y. Moreno, Evaluation of the potential incidence of COVID-19 and effectiveness of containment measures in Spain: a data-driven approach, BMC Med. 18 (2020) 1–12.

[41] C. Yang, J. Wang, A mathematical model for the novel coronavirus epidemic in Wuhan, China, Math. Biosci. Eng. 17 (3) (2020) 2708–2724.

[42] B.F. Maier, D. Brockmann, Effective containment explains subexponential growth in recent confirmed COVID-19 cases in China, Science 368 (6492) (2020) 742–746.

[43] S. Zhao, H. Chen, Modeling the epidemic dynamics and control of COVID-19 outbreak in China, Quant. Biol. 8 (1) (2020) 11–19.

[44] A. Godio, F. Pace, A. Vergnano, SEIR modeling of the Italian epidemic of SARS-CoV-2 using computational swarm intelligence, Int. J. Environ. Res. Public Health 17 (10) (2020) 3535, http://dx.doi.org/10.3390/ijerph17103535, http://dx.doi.org/10.3390/ijerph17103535.

[45] L. Peng, W. Yang, D. Zhang, C. Zhuge, L. Hong, Epidemic analysis of COVID-19 in China by dynamical modeling, 2020, arXiv preprint arXiv:2002.06563.

[46] E. Cheynet, Generalized SEIR epidemic model (fitting and computation), 2020, https://nam03.safelinks.protection.outlook.com/?url=http%3A%2F%2Fgithub.com%2FEcheynet%2FSEIR&data=02%7C01%7Ccorrections.rivervalley%40elsevier.com%7C1dd187211cef4d13e8f008d8348461b8%7C9274ee3f94254109a27f9fb15c10675d%7C0%7C0%7C637317090477039039&sdata=NCXiJurvlJxIvetty87jmbxb%2BKmAwdgzN7RW2GtmtFo%3D&reserved=0.

[47] H. Wan, J.-a. Cui, G.-J. Yang, Risk estimation and prediction by modeling the transmission of the novel coronavirus (COVID-19) in mainland China excluding hubei province, 2020, medRxiv, https://www.medrxiv.org/content/10.1101/2020.03.01.20029629v3.

[48] B. Ivorra, M. Ferrández, M. Vela-Pérez, A. Ramos, Mathematical modeling of the spread of the coronavirus disease 2019 (COVID-19) considering its particular characteristics. The case of China, Commun. Nonlinear Sci. Numer. Simul. 88 (2020) 105303.

[49] T.-M. Chen, J. Rui, Q.-P. Wang, Z.-Y. Zhao, J.-A. Cui, L. Yin, A mathematical model for simulating the phase-based transmissibility of a novel coronavirus, Infect. Dis. Poverty 9 (1) (2020) 1–8.

[50] S.E. Eikenberry, M. Mancuso, E. Iboi, T. Phan, K. Eikenberry, Y. Kuang, E. Kostelich, A.B. Gumel, To mask or not to mask: Modeling the potential for face mask use by the general public to curtail the COVID-19 pandemic, Infect. Dis. Model. 5 (2020) 293–308.

[51] L. López, X. Rodó, The end of social confinement and COVID-19 re-emergence risk, Nat. Hum. Behav. 4 (2020) 746–755.

[52] G. Giordano, F. Blanchini, R. Bruno, P. Colaneri, A. Di Filippo, A. Di Matteo, M. Colaneri, Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy, Nat. Med. 26 (2020) 855–860.

[53] D. Balcan, V. Colizza, B. Gonçalves, H. Hu, J.J. Ramasco, A. Vespignani, Multiscale mobility networks and the spatial spreading of infectious diseases, Proc. Natl. Acad. Sci. 106 (51) (2009) 21484–21489.

[54] D. Balcan, B. Gonçalves, H. Hu, J.J. Ramasco, V. Colizza, A. Vespignani, Modeling the spatial spread of infectious diseases: The global epidemic and mobility computational model, J. Comput. Sci. 1 (3) (2010) 132–145.

[55] D. Balcan, H. Hu, B. Goncalves, P. Bajardi, C. Poletto, J.J. Ramasco, D. Paolotti, N. Perra, M. Tizzoni, W. Van den Broeck, et al., Seasonal transmission potential and activity peaks of the new influenza a (H1N1): a Monte Carlo likelihood analysis based on human mobility, BMC Med. 7 (1) (2009) 45.

[56] M. Chinazzi, J.T. Davis, M. Ajelli, C. Gioannini, M. Litvinova, S. Merler, A.P. y Piontti, K. Mu, L. Rossi, K. Sun, et al., The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak, Science 368 (6489) (2020) 395–400.

[57] R. Li, S. Pei, B. Chen, Y. Song, T. Zhang, W. Yang, J. Shaman, Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2), Science 368 (6490) (2020) 489–493.

[58] A. Aleta, D. Martin-Corral, A.P. y Piontti, M. Ajelli, M. Litvinova, M. Chinazzi, N.E. Dean, M.E. Halloran, I.M. Longini Jr, S. Merler, et al., Modeling the impact of social distancing, testing, contact tracing and household quarantine on second-wave scenarios of the COVID-19 epidemic, 2020, medRxiv, http://dx.doi.org/10.1101/2020.05.06.20092841.

[59] A. Arenas, W. Cota, J. Gomez-Gardenes, S. Gomez, C. Granell, J.T. Matamalas, D. Soriano-Panos, B. Steinegger, Derivation of the effective reproduction number R for COVID-19 in relation to mobility restrictions and confinement, 2020, medRxiv, https://www.medrxiv.org/content/10.1101/2020.04.06.20054320v1.

[60] A. Arenas, W. Cota, J. Gomez-Gardenes, S. Gómez, C. Granell, J.T. Matamalas, D. Soriano-Panos, B. Steinegger, A mathematical model for the spatiotemporal epidemic spreading of COVID19, 2020, MedRxiv, https://www.medrxiv.org/content/10.1101/2020.03.21.20040022v1.

[61] J. Gómez-Gardenes, D. Soriano-Panos, A. Arenas, Critical regimes driven by recurrent mobility patterns of reaction–diffusion processes in networks, Nat. Phys. 14 (4) (2018) 391–395.

[62] D. Soriano-Paños, L. Lotero, A. Arenas, J. Gómez-Gardeñes, Spreading processes in multiplex metapopulations containing different mobility networks, Phys. Rev. X 8 (3) (2018) 031039.

[63] D. Soriano-Paños, J.H. Arias-Castro, A. Reyna-Lara, H.J. Martínez, S. Meloni, J. Gómez-Gardeñes, Vector-borne epidemics driven by human mobility, Phys. Rev. Res. 2 (1) (2020) 013312.

[64] C.R. Wells, P. Sah, S.M. Moghadas, A. Pandey, A. Shoukat, Y. Wang, Z. Wang, L.A. Meyers, B.H. Singer, A.P. Galvani, Impact of international travel and border control measures on the global spread of the novel 2019 coronavirus outbreak, Proc. Natl. Acad. Sci. 117 (13) (2020) 7504–7509.

[65] P. Wang, J.-a. Lu, Y. Jin, M. Zhu, L. Wang, S. Chen, Statistical and network analysis of 1212 COVID-19 patients in Henan, China, Int. J. Infect. Dis. 95 (2020) 391–398.

[66] S. Zhao, Z. Zhuang, P. Cao, J. Ran, D. Gao, Y. Lou, L. Yang, Y. Cai, W. Wang, D. He, et al., Quantifying the association between domestic travel and the exportation of novel coronavirus (2019-nCoV) cases from Wuhan, China in 2020: a correlational analysis, J. Travel Med. 27 (2) (2020) taaa022.

[67] J.O. Lloyd-Smith, S.J. Schreiber, P.E. Kopp, W.M. Getz, Superspreading and the effect of individual variation on disease emergence, Nature 438 (7066) (2005) 355–359.

[68] J. Hellewell, S. Abbott, A. Gimma, N.I. Bosse, C.I. Jarvis, T.W. Russell, J.D. Munday, A.J. Kucharski, W.J. Edmunds, F. Sun, et al., Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts, Lancet Global Health 8 (4) (2020) e488–e496.

[69] M. Perc, N. Gorišek Miksić, M. Slavinec, A. Stožer, Forecasting covid-19, Front. Phys. 8 (2020) 127.

[70] N.M. Ferguson, D.A. Cummings, C. Fraser, J.C. Cajka, P.C. Cooley, D.S. Burke, Strategies for mitigating an influenza pandemic, Nature 442 (7101) (2006) 448–452.

[71] S. Flaxman, S. Mishra, A. Gandy, H. Unwin, H. Coupland, T. Mellan, H. Zhu, T. Berah, J. Eaton, P. Perez Guzman, et al., Report 13: Estimating the number of infections and the impact of non-pharmaceutical interventions on COVID-19 in 11 European countries, 2020.

[72] M. Castro, S. Ares, J.A. Cuesta, S. Manrubia, Predictability: Can the turning point and end of an expanding epidemic be precisely forecast? 2020, arXiv preprint arXiv:2004.08842.
[73] S. Chaudhuri, J.A. Symons, J. Deval, Innovation and trends in the development and approval of antiviral medicines: 1987–2017 and beyond, Antiviral Res. 155 (2018) 76–88.
[74] S. Pushpakom, F. Iorio, P.A. Eyers, K.J. Escott, S. Hopper, A. Wells, A. Doig, T. Guilliams, J. Latimer, C. McNamee, et al., Drug repurposing: progress, challenges and recommendations, Nat. Rev. Drug Discov. 18 (1) (2019) 41–58.
[75] R.A. Hodos, B.A. Kidd, K. Shameer, B.P. Readhead, J.T. Dudley, In silico methods for drug repurposing and pharmacology, Wiley Interdiscip. Rev. Syst. Biol. Med. 8 (3) (2016) 186–210.
[76] N.C. Baker, S. Ekins, A.J. Williams, A. Tropsha, A bibliometric review of drug repurposing, Drug Discov. Today 23 (3) (2018) 661–672.
[77] E. March-Vila, L. Pinzi, N. Sturm, A. Tinivella, O. Engkvist, H. Chen, G. Rastelli, On the integration of in silico drug design methods for drug repurposing, Front. Pharmacol. 8 (2017) 298.
[78] M.J. Keiser, V. Setola, J.J. Irwin, C. Laggner, A.I. Abbas, S.J. Hufeisen, N.H. Jensen, M.B. Kuijer, R.C. Matos, T.B. Tran, et al., Predicting new molecular targets for known drugs, Nature 462 (7270) (2009) 175–181.
[79] B. Mercorelli, G. Palù, A. Loregian, Drug repurposing for viral infectious diseases: how far are we? TIM 26 (10) (2018) 865–876.
[80] M.A. Shereen, S. Khan, A. Kazmi, N. Bashir, R. Siddique, COVID-19 infection: Origin, transmission, and characteristics of human coronaviruses, J. Adv. Res. 24 (2020) 91–98.
[81] I. Astuti, et al., Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2): An overview of viral structure and host response, Diabetes Metab. Syndrome Clin. Res. Rev. 14 (4) (2020) 407–412.
[82] O.O. Glebov, Understanding SARS-CoV-2 endocytosis for COVID-19 drug repurposing, FEBS J. (2020) http://dx.doi.org/10.1111/febs.15369, in press.
[83] L.-s. Wang, Y.-r. Wang, D.-w. Ye, Q.-q. Liu, A review of the 2019 novel coronavirus (COVID-19) based on current evidence, Int. J. Antimicrob. Ag. 55 (6) (2020) 105948.
[84] D.E. Gordon, G.M. Jang, M. Bouhaddou, J. Xu, K. Obernier, K.M. White, M.J. O'Meara, V.V. Rezelj, J.Z. Guo, D.L. Swaney, et al., A SARS-CoV-2 protein interaction map reveals targets for drug repurposing, Nature 583 (2020) 459–468.
[85] H. Zhou, Y. Fang, T. Xu, W.-J. Ni, A.-Z. Shen, X.-M. Meng, Potential therapeutic targets and promising drugs for combating SARS-CoV-2, Br. J. Pharmacol. 177 (2020) 3147–3161.
[86] E. De Clercq, G. Li, Approved antiviral drugs over the past 50 years, Clin. Microbiol. Rev. 29 (3) (2016) 695–747.
[87] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, The protein data bank, Nucleic Acids Res. 28 (1) (2000) 235–242.
[88] L. Zhang, D. Lin, X. Sun, U. Curth, C. Drosten, L. Sauerhering, S. Becker, K. Rox, R. Hilgenfeld, Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors, Science 368 (6489) (2020) 409–412.
[89] W. Rut, Z. Lv, M. Zmudzinski, S. Patchett, D. Nayak, S.J. Snipas, F. El Oualid, M. Bekes, T.T. Huang, M. Drag, et al., Activity profiling and structures of inhibitor-bound SARS-CoV-2-PLpro protease provides a framework for anti-COVID-19 drug design, 2020, bioRxiv, http://dx.doi.org/10.1101/2020.04.29.068890.
[90] Y. Kim, J. Wower, N. Maltseva, C. Chang, R. Jedrzejczak, M. Wilamowski, S. Kang, V. Nicolaescu, G. Randall, K. Michalska, et al., Tipiracil binds to uridine site and inhibits nsp15 endoribonuclease nendou from SARS-CoV-2, 2020, bioRxiv, http://dx.doi.org/10.1101/2020.06.26.173872.
[91] W. Yin, C. Mao, X. Luan, D.-D. Shen, Q. Shen, H. Su, X. Wang, F. Zhou, W. Zhao, M. Gao, et al., Structural basis for inhibition of the RNA-dependent RNA polymerase from SARS-CoV-2 by remdesivir, Science 368 (6498) (2020) 1499–1504.
[92] J. Liu, R. Wang, Classification of current scoring functions, J. Chem. Inform. Model. 55 (3) (2015) 475–482.
[93] Z. Jin, X. Du, Y. Xu, Y. Deng, M. Liu, Y. Zhao, B. Zhang, X. Li, L. Zhang, C. Peng, et al., Structure of m pro from SARS-CoV-2 and discovery of its inhibitors, Nature 582 (2020) 289–293.
[94] A.-T. Ton, F. Gentile, M. Hsing, F. Ban, A. Cherkasov, Rapid identification of potential inhibitors of SARS-CoV-2 main protease by deep docking of 1.3 billion compounds, Mol. Inform. 39 (2020) 2000028.
[95] A. Fischer, M. Sellner, S. Neranjan, M. Smieško, M.A. Lill, Potential inhibitors for novel coronavirus protease identified by virtual screening of 606 million compounds, Int. J. Mol. Sci. 21 (10) (2020) 3626.
[96] B. Tang, F. He, D. Liu, M. Fang, Z. Wu, D. Xu, AI-aided design of novel targeted covalent inhibitors against SARS-CoV-2, 2020, bioRxiv, https://www.biorxiv.org/content/10.1101/2020.03.03.972133v1.full.
[97] S.A. Khan, K. Zia, S. Ashraf, R. Uddin, Z. Ul-Haq, Identification of chymotrypsin-like protease inhibitors of SARS-CoV-2 via integrated computational approach, J. Biomol. Struct. Dyn. (2020) http://dx.doi.org/10.1080/07391102.2020.1751298, in press.
[98] M. Tsuji, Potential anti-SARS-CoV-2 drug candidates identified through virtual screening of the ChEMBL database for compounds that target the main coronavirus protease, FEBS Open Bio 10 (2020) 995–1004.
[99] D.C. Hall Jr, H.-F. Ji, A search for medications to treat COVID-19 via in silico molecular docking models of the SARS-CoV-2 spike glycoprotein and 3CL protease, Travel Med. Infect. Dis. 35 (2020) 101646.
[100] R.S. Joshi, S.S. Jagdale, S.B. Bansode, S.S. Shankar, M.B. Tellis, V.K. Pandya, A. Chugh, A.P. Giri, M.J. Kulkarni, Discovery of potential multi-target-directed ligands by targeting host-specific SARS-CoV-2 structurally conserved main protease, J. Biomol. Struct. Dyn. (2020) http://dx.doi.org/10.1080/07391102.2020.1760137, in press.
[101] D.D. Nguyen, K. Gao, J. Chen, R. Wang, G.-W. Wei, Unveiling the molecular mechanism of SARS-CoV-2 main protease inhibition from 92 crystal structures, 2020, arXiv preprint arXiv:2005.13653.
[102] D.D. Nguyen, K. Gao, M. Wang, G.-W. Wei, MathDL: mathematical deep learning for D3R Grand Challenge 4, J. Comput. Aided Mol. Des. 34 (2) (2020) 131–147.
[103] Z. Cang, G.-W. Wei, Integration of element specific persistent homology and machine learning for protein-ligand binding affinity prediction, Int. J. Numer. Methods Biomed. Eng. 34 (2) (2018) e2914.
[104] B. DasGupta, J. Liang, Models and Algorithms for Biomolecules and Molecular Networks, John Wiley & Sons, 2016.
[105] K. Xia, G.-W. Wei, A review of geometric, topological and graph theory apparatuses for the modeling and analysis of biomolecular data, 2016, arXiv preprint arXiv:1612.01735.
[106] C.S. Pun, K. Xia, S.X. Lee, Persistent-homology-based machine learning and its applications–A survey, 2018, arXiv preprint arXiv:1811.00252.
[107] Z. Cang, G.-W. Wei, Integration of element specific persistent homology and machine learning for protein-ligand binding affinity prediction, Int. J. Numer. Methods Biomed. Eng. 34 (2) (2018) e2914.
[108] E. Estrada, Topological analysis of SARS CoV-2 main protease, Chaos 30 (6) (2020) 061102.
[109] E. Estrada, The Structure of Complex Networks: Theory and Applications, Oxford University Press, 2012.
[110] E. Estrada, G. Silver, Accounting for the role of long walks on networks via a new matrix function, J. Math. Anal. Appl. 449 (2) (2017) 1581–1600.
[111] J. Chen, K. Gao, R. Wang, D.D. Nguyen, G.-W. Wei, Review of COVID-19 antibody therapies, 2020, arXiv preprint arXiv:2006.10584.
[112] L. Abadias, G. Estrada-Rodriguez, E. Estrada, Fractional logarithmic susceptible-infected model. Definition and applications to the study of COVID-19 main protease, Fract. Calc. Appl. Anal. 23 (2) (2020) 635–655.

[113] M. Miotto, L. Di Rienzo, P. Corsi, G. Ruocco, D. Raimondo, E. Milanetti, Simulated epidemics in 3d protein structures to detect functional properties, J. Chem. Inform. Model. 60 (3) (2020) 1884–1891.
[114] M. Rosas-Lemus, G. Minasov, L. Shuvalova, N.L. Inniss, O. Kiryukhina, G. Wiersum, Y. Kim, R. Jedrzejczak, N.I. Maltseva, M. Endres, et al., The crystal structure of nsp10-nsp16 heterodimer from SARS-CoV-2 in complex with S-adenosylmethionine, 2020, bioRxiv, http://dx.doi.org/10.1101/2020.04.17.047498.
[115] Y. Jiang, L. Liu, M. Manning, M. Bonahoom, A. Lotvola, Z.-Q. Yang, Repurposing therapeutics to identify novel inhibitors targeting 2'-O-ribose methyltransferase nsp16 of SARS-CoV-2, 2020.
[116] K. Sharma, S. Morla, A. Goyal, S. Kumar, Computational guided drug repurposing for targeting 2'-O-ribose methyltransferase of SARS-CoV-2, 2020.
[117] R.J. Khan, R.K. Jha, G.M. Amera, M. Jain, E. Singh, A. Pathak, R.P. Singh, J. Muthukumaran, A.K. Singh, Targeting SARS-CoV-2: a systematic drug repurposing approach to identify promising inhibitors against 3C-like proteinase and 2'-O-ribose methyltransferase, J. Biomol. Struct. Dyn. (2020) http://dx.doi.org/10.1080/07391102.2020.1753577, in press.
[118] D. Wrapp, N. Wang, K.S. Corbett, J.A. Goldsmith, C.-L. Hsieh, O. Abiona, B.S. Graham, J.S. McLellan, Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation, Science 367 (6483) (2020) 1260–1263.
[119] M. Shah, B. Ahmad, S. Choi, H.G. Woo, Sequence variation of SARS-CoV-2 spike protein may facilitate stronger interaction with ace2 promoting high infectivity, 2020, http://dx.doi.org/10.21203/rs.3.rs-16932/v1.
[120] O.V. de Oliveira, G.B. Rocha, A.S. Paluch, L.T. Costa, Repurposing approved drugs as inhibitors of SARS-CoV-2 S-protein from molecular modeling and virtual screening, J. Biomol. Struct. Dyn. (2020) http://dx.doi.org/10.1080/07391102.2020.1772885, in press.
[121] L. Caly, J.D. Druce, M.G. Catton, D.A. Jans, K.M. Wagstaff, The FDA-approved drug ivermectin inhibits the replication of SARS-CoV-2 in vitro, Antiviral Res. 178 (2020) 104787.
[122] T. Wei, H. Wang, X. Wu, Y. Lu, S. Guan, F. Dong, C. Dong, G. Zhu, Y. Bao, J. Zhang, et al., In silico screening of potential spike glycoprotein inhibitors of SARS-CoV-2 with drug repurposing strategy, 2020, https://covid-19.conacyt.mx/jspui/handle/1000/4059.
[123] B.O. Villoutreix, J. Creemers, Y. Léger, G. Siegfried, E. Decroly, S. Evrard, A.-M. Khatib, Targeting furin activity through in silico and in vitro drug repurposing strategy for SARS-CoV-2 spike glycoprotein cleavage repression, 2020, http://dx.doi.org/10.21203/rs.3.rs-25856/v1.
[124] M. Örd, I. Faustova, M. Loog, Biochemical evidence of furin specificity and potential for phospho-regulation at spike protein S1/S2 cleavage site in SARS-CoV2 but not in SARS-CoV1 or MERS-CoV, 2020, bioRxiv, http://dx.doi.org/10.1101/2020.04.17.047498.
[125] J.F. Borgio, H.S. Alsuwat, W.M. Al Otaibi, A.M. Ibrahim, N.B. Almandil, L.I. Al Asoom, M. Salahuddin, B. Kamaraj, S. AbdulAzeez, State-of-the-art tools unveil potent drug targets amongst clinically approved drugs to inhibit helicase in SARS-CoV-2, Arch. Med. Sci. 16 (3) (2020) 508.
[126] H.S. Hillen, G. Kokic, L. Farnung, C. Dienemann, D. Tegunov, P. Cramer, Structure of replicating SARS-CoV-2 polymerase, 2020, bioRxiv, http://dx.doi.org/10.1101/2020.04.17.047498.
[127] Z. Ruan, C. Liu, Y. Guo, Z. He, X. Huang, X. Jia, T. Yang, Potential inhibitors targeting RNA-dependent RNA polymerase activity (NSP12) of SARS-CoV-2, 2020, Preprints 2020, 2020030024, http://dx.doi.org/10.20944/preprints202003.0024.v1.
[128] M.U. Mirza, M. Froeyen, Structural elucidation of SARS-CoV-2 vital proteins: Computational methods reveal potential drug candidates against main protease, nsp12 polymerase and nsp13 helicase, J. Pharm. Biomed. Anal. (2020) http://dx.doi.org/10.1016/j.jpha.2020.04.008, in press.
[129] R. Arya, A. Das, V. Prashar, M. Kumar, Potential inhibitors against papain-like protease of novel coronavirus (SARS-CoV-2) from FDA approved drugs, 2020.
[130] C. Wu, Y. Liu, Y. Yang, P. Zhang, W. Zhong, Y. Wang, Q. Wang, Y. Xu, M. Li, X. Li, et al., Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods, Acta Pharm. Sinica B 10 (5) (2020) 766–788.
[131] B.R. Beck, B. Shin, Y. Choi, S. Park, K. Kang, Predicting commercially available antiviral drugs that may act on the novel coronavirus (SARS-CoV-2) through a drug-target interaction deep learning model, Comput. Struct. Biotechnol. J. 18 (2020) 784–790.
[132] A.P. Chiang, A.J. Butte, Systematic evaluation of drug–disease relationships to identify leads for novel drug uses, Clin. Pharmacol. Ther. 86 (5) (2009) 507–510.
[133] M.J. Keiser, B.L. Roth, B.N. Armbruster, P. Ernsberger, J.J. Irwin, B.K. Shoichet, Relating protein pharmacology by ligand chemistry, Nature Biotechnol. 25 (2) (2007) 197–206.
[134] S.B. Smith, W. Dampier, A. Tozeren, J.R. Brown, M. Magid-Slav, Identification of common biological pathways and drug targets across multiple respiratory viruses based on human host gene expression analysis, PLoS One 7 (3) (2012) e33174.
[135] D.M. Gysi, Í.D. Valle, M. Zitnik, A. Ameli, X. Gan, O. Varol, H. Sanchez, R.M. Baron, D. Ghiassian, J. Loscalzo, et al., Network medicine framework for identifying drug repurposing opportunities for covid-19, 2020, arXiv preprint arXiv:2004.07229.
[136] M. Zitnik, M. Agrawal, J. Leskovec, Modeling polypharmacy side effects with graph convolutional networks, Bioinformatics 34 (13) (2018) i457–i466.
[137] Y. Zhou, Y. Hou, J. Shen, Y. Huang, W. Martin, F. Cheng, Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2, Cell Discov. 6 (1) (2020) 1–18.
[138] A. Tahamtan, J. Charostad, S.J. Hoseini Shokouh, M. Barati, An overview of history, evolution, and manufacturing of various generations of vaccines, J. Arch. Mil. Med. 5 (3) (2017) e12315.
[139] W.-H. Chen, U. Strych, P.J. Hotez, M.E. Bottazzi, The SARS-CoV-2 vaccine pipeline: an overview, Curr. Trop. Med. Rep. 7 (2) (2020) 61–64.
[140] J. Zhang, H. Zeng, J. Gu, H. Li, L. Zheng, Q. Zou, Progress and prospects on vaccine development against SARS-CoV-2, Vaccines 8 (2) (2020) 153.
[141] W. Shang, Y. Yang, Y. Rao, X. Rao, The outbreak of SARS-CoV-2 pneumonia calls for viral vaccines, NPJ Vaccines 5 (1) (2020) 1–3.
[142] F. Amanat, F. Krammer, SARS-CoV-2 vaccines: status report, Immunity 52 (4) (2020) 583–589.
[143] M. Ghaebi, A. Osali, H. Valizadeh, L. Roshangar, M. Ahmadi, Vaccine development and therapeutic design for 2019-nCoV/SARS-CoV-2: challenges and chances, J. Cell. Physiol. (2020) http://dx.doi.org/10.1002/jcp.29771, in press.
[144] E. Prompetchara, C. Ketloy, T. Palaga, Immune responses in COVID-19 and potential vaccines: Lessons learned from SARS and MERS epidemic, Asian Pac. J. Allergy Immunol. 38 (1) (2020) 1–9.
[145] C. Palatnik-de Sousa, I. Soares, D. Rosa, Editorial: epitope discovery and synthetic vaccine design, Front. Immunol. 9 (2018) 826.
[146] E. Dorigatti, B. Schubert, Graph-theoretical formulation of the generalized epitope-based vaccine design problem, 2019, bioRxiv, http://dx.doi.org/10.1101/2020.04.17.047498.
[147] B.E. Correia, J.T. Bates, R.J. Loomis, G. Baneyx, C. Carrico, J.G. Jardine, P. Rupert, C. Correnti, O. Kalyuzhniy, V. Vittal, et al., Proof of principle for epitope-focused vaccine design, Nature 507 (7491) (2014) 201–206.
[148] A. Patronov, I. Doytchinova, T-cell epitope vaccine design by immunoinformatics, Open Biol. 3 (1) (2013) 120139.
[149] A.C. Walls, Y.-J. Park, M.A. Tortorici, A. Wall, A.T. McGuire, D. Veesler, Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein, Cell 181 (2) (2020) 281–292.e6.
[150] H. Sivaraman, S.Y. Er, Y.K. Choong, E. Gavor, J. Sivaraman, Structural basis of the SARS-CoV-2/SARS-CoV receptor binding and small-molecule blockers as potential therapeutics, Annu. Rev. Pharmacol. Toxicol. 61 (2021) http://dx.doi.org/10.1146/annurev-pharmtox-061220-093932, in press.

[151] T. Tang, M. Bidon, J.A. Jaimes, G.R. Whittaker, S. Daniel, Coronavirus membrane fusion mechanism offers as a potential target for antiviral development, Antiviral Res. 178 (2020) 104792.
[152] Q. Wang, Y. Zhang, L. Wu, S. Niu, C. Song, Z. Zhang, G. Lu, C. Qiao, Y. Hu, K.-Y. Yuen, et al., Structural and functional basis of SARS-CoV-2 entry by using human ACE2, Cell 181 (4) (2020) 894–904.e9.
[153] M. Hoffmann, H. Kleine-Weber, S. Schroeder, N. Krüger, T. Herrler, S. Erichsen, T.S. Schiergens, G. Herrler, N.-H. Wu, A. Nitsche, et al., SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor, Cell 181 (2) (2020) 271–280.e8.
[154] R. Yan, Y. Zhang, Y. Li, L. Xia, Y. Guo, Q. Zhou, Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2, Science 367 (6485) (2020) 1444–1448.
[155] C.M. Poh, G. Carissimo, B. Wang, S.N. Amrun, C.Y.-P. Lee, R.S.-L. Chee, S.-W. Fong, N.K.-W. Yeo, W.-H. Lee, A. Torres-Ruesta, et al., Two linear epitopes on the SARS-CoV-2 spike protein that elicit neutralising antibodies in COVID-19 patients, Nature Commun. 11 (1) (2020) 1–7.
[156] C.O. Barnes, A.P. West, K. Huey-Tubman, M.A. Hoffmann, N.G. Sharaf, P.R. Hoffman, N. Koranda, H.B. Gristick, C. Gaebler, F. Muecksch, et al., Structures of human antibodies bound to SARS-CoV-2 spike reveal common epitopes and recurrent features of antibodies, Cell 182 (2020) 1–15.
[157] M. Yuan, N.C. Wu, X. Zhu, C.-C.D. Lee, R.T. So, H. Lv, C.K. Mok, I.A. Wilson, A highly conserved cryptic epitope in the receptor binding domains of SARS-CoV-2 and SARS-CoV, Science 368 (6491) (2020) 630–633.
[158] L. Zhang, Multi-epitope vaccines: a promising strategy against tumors and viral infections, Cell. Mol. Immunol. 15 (2) (2018) 182–184.
[159] B. Robson, Computers and viral diseases. preliminary bioinformatics studies on the design of a synthetic vaccine and a preventative peptidomimetic antagonist against the SARS-CoV-2 (2019-nCoV, COVID-19) coronavirus, Comput. Biol. Med. 119 (2020) 103670.
[160] D.W. Kulp, W.R. Schief, Advances in structure-based vaccine design, Curr. Opin. Virol. 3 (3) (2013) 322–331.
[161] G.A. Poland, R.B. Kennedy, B.A. McKinney, I.G. Ovsyannikova, N.D. Lambert, R.M. Jacobson, A.L. Oberg, Vaccinomics, adversomics, and the immune response network theory: individualized vaccinology in the 21st century, in: Seminars in Immunology, Vol. 25, Elsevier, 2013, pp. 89–103.
[162] L. He, J. Zhu, Computational tools for epitope vaccine design and evaluation, Curr. Opin. Virol. 11 (2015) 103–112.
[163] S. Khalili, A. Jahangiri, H. Borna, K. Ahmadi Zanoos, J. Amani, Computational vaccinology and epitope vaccine design by immunoinformatics, Acta Microbiol. Immunol. Hungarica 61 (3) (2014) 285–307.
[164] P. Kalita, A. Padhi, K.Y. Zhang, T. Tripathi, Design of a peptide-based subunit vaccine against novel coronavirus SARS-CoV-2, Microb. Pathog. 145 (2020) 104236.
[165] S. Kang, M. Yang, Z. Hong, L. Zhang, Z. Huang, X. Chen, S. He, Z. Zhou, Z. Zhou, Q. Chen, et al., Crystal structure of SARS-CoV-2 nucleocapsid protein RNA binding domain reveals potential unique drug targeting sites, Acta Pharm. Sinica B (2020) http://dx.doi.org/10.1016/j.apsb.2020.04.009, in press.
[166] M.T. ul Qamar, A. Rehman, U.A. Ashfaq, M.Q. Awan, I. Fatima, F. Shahid, L.-L. Chen, Designing of a next generation multiepitope based vaccine (MEV) against SARS-COV-2: Immunoinformatics and in silico approaches, 2020, BioRxiv, http://dx.doi.org/10.1101/2020.02.28.970343.
[167] M. Bhattacharya, A.R. Sharma, P. Patra, P. Ghosh, G. Sharma, B.C. Patra, S.-S. Lee, C. Chakraborty, Development of epitope-based peptide vaccine against novel coronavirus 2019 (SARS-COV-2): Immunoinformatics approach, J. Med. Virol. 92 (6) (2020) 618–631.
[168] A. Grifoni, J. Sidney, Y. Zhang, R.H. Scheuermann, B. Peters, A. Sette, A sequence homology and bioinformatic approach can predict candidate targets for immune responses to SARS-CoV-2, Cell Host Microbe 27 (64) (2020) 671–680.e2.
[169] M. Zheng, L. Song, Novel antibody epitopes dominate the antigenicity of spike glycoprotein in SARS-CoV-2 compared to SARS-CoV, Cell. Mol. Immunol. 17 (5) (2020) 536–538.
[170] K.M. Campbell, G. Steiner, D.K. Wells, A. Ribas, A. Kalbasi, Prediction of SARS-CoV-2 epitopes across 9360 HLA class I alleles, 2020, BioRxiv, http://dx.doi.org/10.1101/2020.03.30.016931.
[171] K. Kiyotani, Y. Toyoshima, K. Nemoto, Y. Nakamura, Bioinformatic prediction of potential T cell epitopes for SARS-CoV-2, J. Hum. Genet. 65 (7) (2020) 569–575.
[172] M. Prachar, S. Justesen, D.B. Steen-Jensen, S.P. Thorgrimsen, E. Jurgons, O. Winther, F.O. Bagger, Covid-19 vaccine candidates: Prediction and validation of 174 SARS-CoV-2 epitopes, 2020, bioRxiv, http://dx.doi.org/10.1101/2020.04.17.047498.
[173] S.F. Ahmed, A.A. Quadeer, M.R. McKay, Preliminary identification of potential vaccine targets for the COVID-19 coronavirus (SARS-CoV-2) based on SARS-CoV immunological studies, Viruses 12 (3) (2020) 254.
[174] S. Cornet, I. Miconnet, J. Menez, F. Lemonnier, K. Kosmatopoulos, Optimal organization of a polypeptide-based candidate cancer vaccine composed of cryptic tumor peptides with enhanced immunogenicity, Vaccine 24 (12) (2006) 2102–2109.
[175] B. Schubert, O. Kohlbacher, Designing string-of-beads vaccines with optimal spacers, Genome Med. 8 (1) (2016) 9.
[176] N.C. Toussaint, Y. Maman, O. Kohlbacher, Y. Louzoun, Universal peptide vaccines–optimal peptide vaccine design based on viral sequence conservation, Vaccine 29 (47) (2011) 8745–8753.
[177] J. Theiler, B. Korber, Graph-based optimization of epitope coverage for vaccine antigen design, Stat. Med. 37 (2) (2018) 181–194.
[178] E. Dorigatti, B. Schubert, Joint epitope selection and spacer design for string-of-beads vaccines, 2020, bioRxiv, http://dx.doi.org/10.1101/2020.04.17.047498.
[179] T. Vider-Shalit, S. Raffaeli, Y. Louzoun, Virus-epitope vaccine design: informatic matching the HLA-I polymorphism to the virus genome, Mol. Immunol. 44 (6) (2007) 1253–1261.
[180] L. Martínez, M. Milanič, I. Malaina, C. Álvarez, M.-B. Pérez, I. M. de la Fuente, Weighted lambda superstrings applied to vaccine design, PLoS One 14 (2) (2019) e0211714.
[181] L. Martínez, M. Milanič, L. Legarreta, P. Medvedev, I. Malaina, M. Ildefonso, A combinatorial approach to the design of vaccines, J. Math. Biol. 70 (6) (2015) 1327–1358.
[182] S. Eubank, I. Eckstrand, B. Lewis, S. Venkatramanan, M. Marathe, C. Barrett, Commentary on Ferguson, et al., "Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand", Bull. Math. Biol. 82 (2020) 1–7.
[183] S. Flaxman, S. Mishra, A. Gandy, H.J.T. Unwin, T.A. Mellan, H. Coupland, C. Whittaker, H. Zhu, T. Berah, J.W. Eaton, et al., Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe, Nature (2020) http://dx.doi.org/10.1038/s41586-020-2405-7, in press.